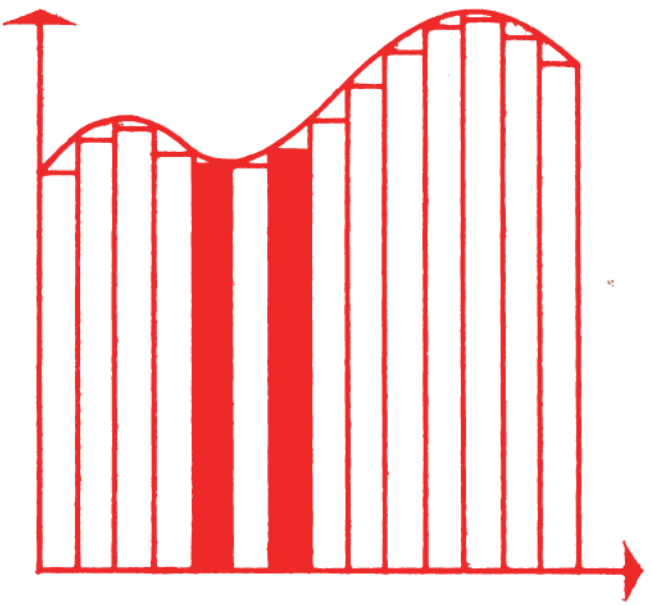
D.K. Faddeev, M.S. Nikulin, and I.F. Sokolovsky

Elements of higher Mathematics

For High-School Students



Mir Publishers Moscow

About the Book

This book considers the principal concepts of the differential and integral calculus, their applications in the investigation of elementary functions, their use in approximate calculations and in the solution of problems in mechanics and physics. Separate chapters are devoted to the study of trigonometric functions, complex numbers and elements of the probability theory. Each chapter is supplied with exercises.

The book is intended for high-school students, students of technical colleges, teachers of mathematics, engineers and technicians.

ISBN 5-03-000535-8

Д. К. Фаддеев, М. С. Никулин, И. Ф. Соколовский

ЭЛЕМЕНТЫ ВЫСШЕЙ МАТЕМАТИКИ ДЛЯ ШКОЛЬНИКОВ

Москва «Наука»

D.K. Faddeev, M.S. Nikulin,
and I.F. Sokolovsky

Elements of higher Mathematics

For High-School Students

Mir Publishers Moscow

Translated from the Russian by Irene Aleksanova

First published 1989
Revised from the 1987 Russian edition

На английском языке

Printed in the Union of Soviet Socialist Republics

ISBN 5–03000535–8

© Издательство «Наука. Главная редакция физико-математической литературы, 1987
© English translation, Mir Publishers, 1989

PREFACE

The main purpose of the book is to show that the principal ideas of mathematical analysis are simple and graphic when they are presented at the same intuitive level at which they actually originated. Therefore in the first pages of the book we introduce the simple notion which we call the main principle of differential calculus. According to it, a sufficiently small segment of a smooth curve almost coincides with a segment of a straight line and the difference between them gradually disappears as the segment of the curve converges to a point. Wherever possible, strict proofs are preceded by arguments based on this principle.

The proofs of theorems given in this book are more complete and more strict than those usually given in school textbooks, but this was not our main task. These places in the book and also the places marked by the symbols ◀ and ▶ can be omitted in the course of the first reading.

We think that the understanding of this main principle of differential calculus and the nonformal application of this principle have a great methodological and philosophic meaning and must precede the study of serious university courses in mathematical analysis.

We wish to express our gratitude to school teachers of Leningrad N.E. Skuchas and S.A. Ioffe who used the material presented in this book to teach algebra and fundamentals of analysis at school. We hope that this book will be of use for high-school students interested in mathematics and its applications in physics and engineering, for students of technical colleges and higher schools and for teachers of mathematics and physics.

CONTENTS

Preface 5

Chapter 1. Principal Concepts of Differential Calculus 10

1.1. The Main Principle of Differential Calculus 10
1.2. Infinitely Small Quantities 14
1.3. Convergent Variables and Their Limits 15
1.4. Infinitely Large Quantities 18
1.5. Examples of Calculating Limits 18
1.6. Limits of Functions 19
1.7. Continuity of Functions 21
1.8. A More Exact Definition of the Concept of a Derivative 27
1.9. The Equation of a Tangent to the Graph of a Function 30
1.10. The Rate of Variation of a Function 31
1.11. The Velocity of Mechanical Motion of a Particle in a Straight Line 33
1.12. The Differential of a Function 37
1.13. The Differential of a Function of a Function 41
Exercises to Chapter 1 43

Chapter 2. Technique of Differentiation 51

2.1. Differentiation of the Results of Arithmetic Operations 51
2.2. Differentiation of a Logarithmic Function 57
2.3. Proof of the Existence of a Limit of the Function $(1 + h)^{1/h}$ as $h \to 0$ 59
2.4. Differentiation of an Exponential Function 63
2.5. Differentiation of a Power Function 64
2.6. Differentiation of Functions Defined by Equations 65
Exercises to Chapter 2 66

Chapter 3. Some Applications of Differential Calculus 70

3.1. Tests for Increase and Decrease of Functions 70
3.2. Refinement of the Proofs of the Theorems on the Increase and Decrease of Functions 74
3.3. Maxima and Minima of Functions 79
3.4. A Simple Example and Some Inferences from Its Consideration 87
3.5. Higher-Order Derivatives 90
3.6. The Binomial Theorem 90

3.7. Investigation of Functions with the Use of Higher-Order Derivatives 93
3.8. The Order of Smallness of a Function in the Neighbourhood of a Point at Which the Function Vanishes and the Proximity of Functions 98
3.9. The Relationship Between the Order of Smallness and the Order of the First Nonzero Derivative 100
3.10. Formulas of Taylor and Maclaurin 105
3.11. General Concepts of the Theory of Approximation 109
3.12. Estimation of the Errors of the Results of Calculations Involving Approximate Numbers 111
Exercises to Chapter 3 116

Chapter 4. Trigonometric Functions 128

4.1. Generalization of the Concept of an Angle 128
4.2. Measuring Angles in Radians 130
4.3. The Sine and Cosine Functions 132
4.4. The Fundamental Properties of the Sine and Cosine Functions 135
4.5. Reducing the Values of the Sine and Cosine Functions to the Values on the Interval $0 \leqslant \varphi \leqslant \pi/4$ 140
4.6. The Tangent and Cotangent Functions 140
4.7. Expressing Trigonometric Functions in Terms of One Another 145
4.8. A Significant Limit 146
4.9. The Graphs of the Functions $y = \sin x$ and $y = \cos x$ 147
4.10. The Graphs of the Functions $y = \tan x$ and $y = \cot x$ 151
Exercises to Chapter 4 152

Chapter 5. Transformation of Expressions Which Include Trigonometric Functions. Some Applications 166

5.1. The Sine and Cosine of the Sum and the Difference of Arguments 166
5.2. The Tangent and Cotangent of a Sum and a Difference 169
5.3. Trigonometric Functions of a Doubled Argument and of Some Multiple Arguments 171
5.4. Trigonometric Functions of a Half-Argument 174
5.5. Expressing Trigonometric Functions in Terms of the Tangent of a Half-Argument 175
5.6. Expressing the Products of the Sine and Cosine Functions as Sums and Expressing Sums as Products 176
5.7. Transformation of a Linear Combination of a Sine and a Cosine 178
5.8. Harmonic Oscillations 179
5.9. Oscillations with a Variable Amplitude 181
5.10. Basic Trigonometric Equations and Inverse Trigonometric Functions 183

5.11. Some Actions on Direct and Inverse Trigonometric Functions 189
5.12. Trigonometric Equations 191
5.13. Solving Basic Trigonometric Inequalities 200
5.14. Trigonometric Inequalities of a More General Form 204
5.15. Sample Proofs of Inequalities Involving Trigonometric Expressions 210
5.16. Differentials and Derivatives of Trigonometric Functions 212
5.17. Application of Derivatives to the Investigation of Functions Which Can be Expressed in Terms of Trigonometric Functions 216
5.18. Derivatives and Differentials of Inverse Trigonometric Functions 224
Exercises to Chapter 5 226

Chapter 6. Elements of Integral Calculus 236
6.1. Definition of Integration 236
6.2. A More Rigorous Proof of the Lemma 237
6.3. Basic Integration Formulas 238
6.4. Integration Based on the Use of the Invariance of the Formula for the Differential of a Function of a Function 241
6.5. Integration by Parts 243
6.6. The Area of a Curvilinear Trapezoid 244
6.7. Basic Properties of the Definite Integral 252
6.8. Representing an Integral as a Sum 256
6.9. The Integral as the Limit of a Sum 257
6.10. Approximate Calculation of Integrals 263
6.11. The Volume of a Solid of Revolution 268
6.12. The Length of the Arc of a Curve 270
6.13. The Lateral Area of a Solid of Revolution 273
6.14. The Differential Equation 274
6.15. Some Differential Equations Important for Mechanics 276
Exercises to Chapter 6 280

Chapter 7. Complex Numbers 298
7.1. Introductory 298
7.2. Basic Definitions 301
7.3. Trigonometric Form of a Complex Number 304
7.4. Multiplication of Complex Numbers in Trigonometric Form 307
7.5. Extracting a Root of a Complex Number 309
7.6. Extracting a Square Root of a Complex Number 312
7.7. An Exponential and a Logarithmic Function of a Complex Variable 314
Exercises to Chapter 7 317

Chapter 8. Elements of Combinatorics and the Probability Theory 320

8.1. Simple Combinatorics Problems 320
8.2. On Probabilities 328
8.3. Addition of Probabilities 333
8.4. Multiplication of Probabilities 334
8.5. Application of Probabilities to Genetics 337
8.6. Random Variables 340
8.7. The Sum of Independent Random Variables 342
8.8. Expectation and Variance of a Random Variable Connected with Bernoulli Trials 344
8.9. Chebyshev's Inequality 346
8.10. The Law of Large Numbers for the Bernoulli Trials 346
8.11. Random Walk in the Plane 347
8.12. Random Variables Whose Values Are Concentrated in an Interval or Throughout the Real Axis 349
8.13. Buffon's Needle Problem 354

Hints and Answers 357

Index 364

Chapter 1

PRINCIPAL CONCEPTS OF DIFFERENTIAL CALCULUS

1.1. The Main Principle of Differential Calculus

The main purpose of the branch of mathematics known as mathematical analysis is to investigate functions which are "good" enough, namely, which originate in the course of quantitative investigation of problems of classical physics, mechanics and technology. Of considerable significance is the study of functions "in the small", i.e. *in a sufficiently small neighbourhood of every point* of the domain of definition of the function. The most simple and important problems which require an answer are whether the function increases (decreases) in the neighbourhood of the point and what is the rate of this increase (decrease). The study of functions "in the small" is the subject of the division of analysis known as *differential calculus.*

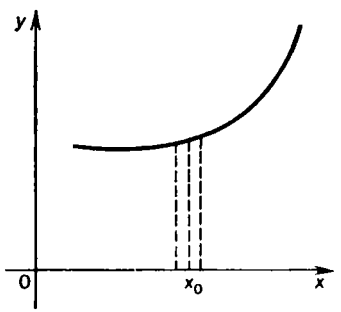


Fig. 1.1

Of no less importance is the inverse problem, i.e. how to reconstruct the function in the whole, i.e. in the entire domain of its definition if the behaviour of the function in the neighbourhood of every point of its domain is known. This problem is the subject of the study of the so-called *integral calculus.*

The *main principle* of differential calculus is very simple. Essentially it consists in the following. Assume that we are given a smooth function $y = f(x)$, i.e. a function whose graph is a continuous smooth curve without sharp bends. We take a small interval adjoining a given point x_0 (Fig. 1.1). If the interval we have chosen is really small, then the graph of the function will not have time to curve and will approach a segment of a straight line. If we make the interval smaller and smaller, then the segment of the graph will still closer

approach the segment of the straight line. The portion of the graph will approach the straight line nearer and nearer. A straight line is the graph of a linear function. Consequently, *we can say that any smooth function "in the small", i.e. on a sufficiently small interval of variation of the argument, is almost linear and, the smaller the interval, the closer the given function is to a linear function.*

This is the **gist** of the main principle of differential calculus.

It stands to reason that the formulation of this principle **is not a theorem in the strict sense of the word** since both the assumption (the smoothness of the function) and the inference (the closeness of the function to a linear function) are given here by visual geometric representation rather than by strict mathematical definitions, and this is the drawback of the formulation, but its strength is in its simplicity and visuality, so it can be easily used in applications.

The straight line to which the graph of the function $y = f(x)$ adjoins most closely in the neighbourhood of the point with abscissa x_0 is a *tangent* to the graph at this point. Its slope is the *value of the derivative* of the function $f(x)$ at the point x_0. The value of the derivative at the point x_0 is designated as $f'(x_0)$ or $y'_{x=x_0}$.

The value of the derivative $f'(x_0)$ depends on the abscissa x_0 of the point for which this value is being calculated. Therefore $f'(x_0)$ is the value of a function known as a *derivative of the function* $f(x)$. This function is designated as $f'(x)$.

Let us consider some examples showing how a smooth function approaches a linear function on a small interval.

Example 1. Let $y = f(x) = x^2$ and $x_0 = 1$. In the vicinity of the point $x_0 = 1$ we can set $x = 1 + h$, where h is a variable which assumes arbitrarily small values. Then

$$x^2 = (1+h)^2 = 1+2h+h^2 \cong 1+2h = 1+2(x-1) = 2x-1.$$

When passing to the approximate equality, we have discarded h^2 taking into account that $h^2 = (x-1)^2$ diminishes essentially faster than $|h| = |x-1|$ since h^2 is as many times smaller than $|h|$ as $|h|$ is smaller than 1.

Thus the ordinate of a point on the graph of the function $f(x) = x^2$ differs from the corresponding ordinate on the straight line $y = 2x - 1$ by the quantity $h^2 = |x-1|^2$ which

is essentially smaller for small $|h|$ than the deviation $|h| = |x - 1|$ of the abscissa x from 1.

Thus the straight line defined by the equation $y = 2x - 1$ is especially close to the graph of the function $y = x^2$ in the neighbourhood of the point $x_0 = 1$, i.e. this straight line is a tangent to the graph at this point and its slope 2 is the value $f'(1)$ of the derivative for $x_0 = 1$.

▲ It is easy to make sure that the straight line $y = 2x - 1$ is closer to the graph of the function x^2 at the point (1,1) than any other straight line which passes through the point (1, 1). The equation of any oblique straight line which passes through the point (1,1) is $y = kx - (k - 1)$, where k is its slope. The ordinate of the point on the graph of the function x^2 differs from the ordinate of the corresponding point on the straight line by the quantity

$$x^2 - (kx - (k - 1)) = x^2 - kx + k - 1 = (x - 1)(x - k + 1).$$

This quantity is as many times smaller in absolute value than $|x - 1|$ as $|x - k + 1| = |x - 1 + 2 - k|$ is smaller than 1. But when $k \neq 2$, for x sufficiently close to 1 the quantity $|x - 1 + 2 - k|$ becomes arbitrarily close to $|2 - k| \neq 0$ whereas when $k = 2$, the quantity $|x - 1 + 2 - k| = |x - 1|$ becomes arbitrarily small. Thus the difference $x^2 - (kx - (k - 1))$ becomes the smallest particularly for $k = 2$ for sufficiently small $|x - 1|$. ▶

We shall now find the value of the derivative of the function $f(x) = x^2$ at a point with abscissa x_0. To do this, we shall calculate the values of the function for x which is close to x_0, for which purpose we set $x = x_0 + h$ and assume that $|h|$ decreases indefinitely. We have

$$\begin{aligned} x^2 = (x_0 + h)^2 &= x_0^2 + 2x_0h + h^2 \cong x_0^2 + 2x_0h \\ &= x_0^2 + 2x_0(x - x_0) = 2x_0x - x_0^2. \end{aligned}$$

When passing to the approximate equality, we again discarded $h^2 = (x - x_0)^2$ as a quantity which decreases essentially faster than $|h|$. Therefore the linear function $2x_0x - x_0^2$ has a tangent to the graph of the function x^2 at the point x_0 as its graph. Its slope $2x_0$ is the value of the derivative of the function x^2 at the point x_0. Therefore, this derivative, regarded as a function, is equal to $2x$. Thus $(x^2)' = 2x$.

Exercises

◀ 1. Find a straight line which comes to the graph of the function $y = \sqrt{x}$ in the vicinity of the point $x = x_0$ closer than all other straight lines.

The equation of the straight line which passes through the point $(x_0, \sqrt{x_0})$ has the form $y = k(x - x_0) + \sqrt{x_0}$. Let $x = x_0 + h$, where h is a variable assuming arbitrarily small values (tending to zero). We consider the difference of the functions

$$\sqrt{x} \text{ and } k(x-x_0)+\sqrt{x_0}$$

in the vicinity of the point $x = x_0$:

$$\begin{aligned}\sqrt{x_0+h}-k(x_0+h-x_0)-\sqrt{x_0} &= \sqrt{x_0+h}-\sqrt{x_0}-kh\\ &= \frac{x_0+h-x_0}{\sqrt{x_0+h}+\sqrt{x_0}}-kh\\ &= h\left(\frac{1}{\sqrt{x_0+h}+\sqrt{x_0}}-k\right).\end{aligned}$$

It is clear that if $k=\frac{1}{2\sqrt{x_0}}$, then, for h tending to zero, not only the first but also the second factor $\left(\frac{1}{\sqrt{x_0+h}+\sqrt{x_0}}-\frac{1}{2\sqrt{x_0}}\right)$ tends to zero indefinitely. If $k \neq \frac{1}{2\sqrt{x_0}}$, then the absolute value of the second factor tends to assume a nonzero value as h tends to zero indefinitely. This means that the straight line with the slope $k = \frac{1}{2\sqrt{x_0}}$ is closer to the graph of the function $y=\sqrt{x}$ in the vicinity of the point $x=x_0$ than any other straight line. According to the definition, the number $\frac{1}{2\sqrt{x_0}}$ is the value of the derivative of the function $\sqrt{x}$ at the point $x=x_0$ and the function $y = \frac{1}{2\sqrt{x}}$ is the derivative of the function $y=\sqrt{x}$. ▶

2. Calculate the value of the derivative of the function $y = \sqrt{x}$ at the point $x = 4$. Write the equation of the tangent to the graph of the function $y = \sqrt{x}$ at the point with abscissa $x = 4$.

3. Use a computer to tabulate the values of the functions $y = \sqrt{x}$ and $y = \frac{1}{4}x + 1$ on the interval [3.95, 4.05] with a stepsize 0.01. Use the table to construct the graphs of these functions on the interval [3.95, 4.05] assuming 50 cm to be the scale unit (do not try to draw

the coordinate axes in the figure). Make sure that with this accuracy of construction the graphs of the functions $y = \sqrt{x}$ and $y = \frac{1}{4}x + 1$ are indistinguishable.

1.2. Infinitely Small Quantities

When considering functions "in the small" we have to deal with variables whose values become arbitrarily close to zero (Sec. 1.1). Variables of this kind are known as vanishing, or *infinitely small (infinitesimal) quantities.* Pay attention to the fact that an **infinitely small quantity is a variable quantity,** and not a constant, which occupies an "intermediate" position between zero and other numbers.

Here we give a definition of an infinitely small quantity and some of its properties which are useful for our discussion. We shall only consider a variable which assumes an infinite sequence of values dependent on the number n. Recall that the sequence $x_1, x_2, \ldots, x_n, \ldots$ of the values of the variable x is *vanishing (infinitely decreasing) if all of its terms, beginning with a certain number, become arbitrarily small in absolute value.* For example, the terms of the sequence $1, \frac{1}{2}, \frac{1}{3}, \ldots, \frac{1}{n}, \ldots$ become smaller than 0.01 beginning with the 101st term, become smaller than 10^{-6} beginning with the 1 000 001st term, etc. The variable x itself, which assumes a vanishing sequence of values, is a *vanishing*, or *infinitely small (infinitesimal) quantity.*

Let us consider the results of arithmetic operations involving infinitesimals.

Theorem 1. *If x is a vanishing variable and a is a constant, then their product ax is also a vanishing variable.*

Proof. If $a = 0$, then ax assumes only a zero value. According to the definition, we must consider such a zero "variable" to be vanishing (even to have already "vanished"). Now if $a \neq 0$, then for any $c > 0$ the values $|ax_n|$ of the variable $|ax|$ become smaller than c as soon as $|x_n|$ becomes smaller than $\frac{c}{|a|}$, and this occurs beginning with a sufficiently large n.

Theorem 2. *If x and y are vanishing variables, then their sum $x + y$ is also vanishing.*

Proof. Indeed, $| x_n + y_n |$ becomes smaller than the given $c > 0$ as soon as $| x_n |$ and $| y_n |$ become smaller than $c/2$, and this occurs beginning with a sufficiently large n since x and y are vanishing variables.

The following theorem is still more obvious.

Theorem 3. *The product xy of two vanishing variables x and y is a vanishing variable.*

Proof. Indeed, $| x_n y_n |$ is as many times smaller than $| x_n |$ as $| y_n |$ is smaller than 1.

We say that the vanishing variable z approaches zero *essentially faster* than the vanishing variable x if there is a vanishing variable y such that $z = xy$. The values z_n/x_n are the values of the variable y for n for which $x_n \neq 0$ (for $x_n = 0$ the value z_n must be equal to zero.). In this case we also say that z is an infinitesimal of a higher order of smallness than x. Thus, for instance, the variable z with values $z_n = 2/n^2$ approaches zero essentially faster than the variable x with values $x_n = 1/n$. Indeed, here y has values $y_n = 2/n$ which tend to zero indefinitely.

It is clear that the product of two infinitesimals approaches zero essentially faster than each of the cofactors.

A variable is *bounded* if all of its values are smaller in absolute value than some positive number a, or, to put it otherwise, if all the values are in the interval $(-a, a)$ for a certain $a > 0$.

Theorem 4. *The product xy of the vanishing variable x by a bounded variable y is a vanishing variable.*

Proof. $| x_n y_n | = |x_n | \cdot |y_n | < | x_n | a$, where a is a number bounding the moduli of the values of the variable y. The values $| x_n | a$ become arbitrarily small beginning with a certain n, and this is obviously also true for $| x_n y_n |$ which are smaller than $| x_n | a$.

Exercises

1. Show that the variable x which assumes a sequence of values $x_n = a^n$, where $a = \text{const}$ and $0 < a < 1$, is a vanishing variable.

2. Show that if $0 < a < b < 1$, then the sequence $x_n = a^n$ approaches zero essentially faster than the sequence $y_n = b^n$.

1.3. Convergent Variables and Their Limits

Recall that the sequence $x_1, x_2, \ldots, x_n, \ldots$ is *convergent* if, beginning with a certain n, all of its terms come arbitrarily close to a definite number a. This number is the *limit*

of the sequence $x_1, \ldots, x_n, \ldots$ and is designated as $\lim\limits_{n\to\infty} x_n$, or $\lim x_n$, as $n \to \infty$. When considering sequences, we shall often write $\lim x_n$. If the variable x assumes the convergent sequence of values $x_1, x_2, \ldots, x_n, \ldots$, then the variable x is *convergent*. The limit of the sequence of its values is its *limit*. The fact that the number a is the limit of the variable x is written in the form of the formula $\lim x = a$, or $x \to a$ (the latter notation must be read as "x tends to a").

It is clear that any infinitely small variable converges and 0 is its limit. We often write $x \to 0$ or "x tends to 0" instead of writing "x is an infinitesimal".

We can see from the definition of the limit of a convergent variable that *if* $x \to a$, *then* $x - a$ *is an infinitesimal.* The converse is also true; *if* $x - a = z$, i.e. $x = a + z$, *where* a *is a constant and* z *is an infinitesimal, then* x *is a convergent variable and* a *is its limit.*

Theorem 5. *If* x *and* y *are convergent variables with limits* a *and* b *respectively, then their sum and product also converge to the limits* $a + b$ *and* ab *respectively.*

Proof. $x = a + z$, $y = b + t$, where z and t are infinitesimals. Then $x + y = (a + b) + (z + t)$ and $xy = ab + (at + bz + zt)$. The second terms on the right-hand sides of these relations are infinitely small quantities. Consequently, $a + b = \lim (x + y)$ and $ab = \lim (xy)$.

Theorem 6. *If the variable* x *has a nonzero limit* a, *then* $1/x$ *has a limit* $1/a$.

Proof. $\frac{1}{x} - \frac{1}{a} = \frac{a - x}{ax} = -\frac{1}{a} \cdot \frac{1}{x} (x - a)$. The factor $1/x$ is bounded beginning with a certain place: sufficiently distant values x_n of the variable x will become larger in absolute value than $\frac{2}{3} \mid a \mid$. For this to occur, it is only necessary that x_n should differ from a less than by $\frac{1}{3} \mid a \mid$. Therefore, beginning with this place, $\left|\frac{1}{x_n}\right| < \frac{3}{2 \mid a \mid}$. Furthermore, in the expression $-\frac{1}{a} \cdot \frac{1}{x} (x - a)$ the factor $-1/a$ is constant and the factor $x - a$ is an infinitesimal. Consequently, $\frac{1}{x} - \frac{1}{a} =$

$\frac{-1}{a}\cdot\frac{1}{x}(x-a)$ is also an infinitesimal and $1/x$ converges to $1/a$.

Theorem 7. *If* $\lim y = b$ *and* $\lim x = a \neq 0$, *then* y/x *has a limit and this limit is equal to* b/a.

Proof. $y/x = y\cdot\frac{1}{x}$. The limit of the first factor is b and, by virtue of Theorem 6, the limit of the second is $1/a$. According to Theorem 5, the limit of $\frac{y}{x} = y\cdot\frac{1}{x}$ exists and is equal to $b\cdot\frac{1}{a} = \frac{b}{a}$.

Example 1. Employing the definition of the limit of a variable quantity, show that the variable quantity x, which assumes a sequence of values $x_n = \frac{5n}{n+2}$, has the number 5 as its limit.

Solution. Let us show that beginning with a certain number all the terms of the sequence x_n approach the number 5 arbitrarily close.

The distance between x_n and the number 5 is

$$|x_n - 5| = \left|\frac{5n}{n+2} - 5\right| = \left|\frac{-10}{n+2}\right| = \frac{10}{n+2} \to 0.$$

◀ Indeed, let ε be an arbitrary positive number. Solving the inequality

$$|x_n - 5| = \frac{10}{n+2} < \varepsilon,$$

we obtain $10 < \varepsilon(n+2)$, $n > \frac{10}{\varepsilon} - 2$.

Thus all the terms of the sequence x_n, beginning with the number larger than $\frac{10}{\varepsilon} - 2$, are at a distance smaller than ε, from the number 5. Since we can assume ε to be an arbitrarily small number, this means that all the terms of the sequence, beginning with a certain number, approach the number 5 arbitrarily close, i.e. the number 5 is the limit of the sequence x_n: $\lim x_n = 5$, and this means that 5 is the limit of the variable x: $\lim x = 5$. ▶

Using the concept of an infinitely small variable, we can substantiate the statement that $\lim x = 5$ as follows. The difference $x - 5$ is equal to $\frac{5n}{n+2} - 5 = -\frac{10}{n+2}$. It is clear that as the number n increases indefinitely, the fraction

$-\frac{10}{n+2}$ tends to zero indefinitely, i.e. is an infinitely small (vanishing) variable and, consequently, the number 5 is the limit of the variable x.

1.4. Infinitely Large Quantities

The variable x, whose values become, beginning with a certain number, arbitrarily large in absolute value, is said to be *infinitely large* or *tending to infinity*. This is written as $x \to \infty$ or $\lim x = \infty$ (one must not understand these notations as meaning that there is a "number" infinity ∞ to which the values of the variable tend indefinitely). We can cite a variable with values $x_n = n$ running through a series of natural numbers in a natural order as a typical example of an infinitely large variable.

Sometimes it is expedient to emphasize in the notation the fact that the values of the variable become arbitrarily large positive numbers. This is written as $x \to +\infty$ or $\lim x = +\infty$. The notations $x \to -\infty$ and $\lim x = -\infty$ mean that the values of x become arbitrarily large in absolute value, remaining negative.

If y is a vanishing quantity which does not assume the value zero, then the variable $x = 1/y$ is evidently infinitely large since the inverses of infinitesimals become arbitrarily large.

1.5. Examples of Calculating Limits

Example 1. Given $x_n \to 2$, find $\lim\limits_{n\to\infty} \frac{x_n^2+x_n+2}{x_n+2}$.

By virtue of Theorem 5, the limit of x_n^2 is $2^2 = 4$, the limit of $x_n^2 + x_n + 2$ is $4 + 2 + 2 = 8$, the limit of $x_n + 2$ is $2 + 2 = 4$. By virtue of Theorem 7, the required limit is $8/4 = 2$.

Example 2. Find $\lim\limits_{n\to\infty} \frac{n+1}{2n+1}$.

We put n before the brackets in the numerator and the denominator and cancel the fraction. We obtain

$$\frac{n+1}{2n+1} = \frac{1+(1/n)}{2+(1/n)}.$$

As $n \to \infty$, the quantity $1/n$ is vanishing so that the limit of the quantity $1 + (1/n)$ is 1, the limit of $2 + (1/n)$ is 2. Applying Theorem 7, we get

$$\lim_{n\to\infty} \frac{n+1}{2n+1} = \lim_{n\to\infty} \frac{1+(1/n)}{2+(1/n)} = \frac{1}{2}.$$

Example 3. Find $\lim\limits_{n\to+\infty} (\sqrt{n^2+n} - n)$. We have to find the limit of the difference of the square roots $\sqrt{n^2+n}$ and $n = \sqrt{n^2}$ of rather close numbers.

We set

$$\sqrt{n^2+n} - n = \frac{(\sqrt{n^2+n} - n)(\sqrt{n^2+n} + n)}{\sqrt{n^2+n} + n} = \frac{n^2+n-n^2}{\sqrt{n^2+n} + n}$$

$$= \frac{n}{n\left(\sqrt{1+\frac{1}{n}} + 1\right)} = \frac{1}{\sqrt{1+\frac{1}{n}} + 1}.$$

It is clear that $\sqrt{1 + \frac{1}{n}} \to 1$ as $n \to \infty$. Therefore

$$\lim_{n\to\infty} (\sqrt{n^2+n} - n) = \lim_{n\to\infty} \frac{1}{\sqrt{1+\frac{1}{n}} + 1} = \frac{1}{1+1} = \frac{1}{2}.$$

1.6. Limits of Functions

Assume that the function $y = f(x)$ is defined at all points of an interval except, maybe, a point x_0 which is within the interval or on its boundary. We say that the number b is the limit of the function $f(x)$ as $x \to x_0$ if for any sequence of values $x_1, x_2, \ldots, x_n, \ldots$ of the variable which belongs to this interval and converges to x_0 ($x_n \neq x_0$) the sequence of values $f(x_1), f(x_2), \ldots, f(x_n), \ldots$ of the function converges to b. This is written as

$$\lim_{x\to x_0} f(x) = b.$$

The requirement that $\lim\limits_{x_n\to x_0} f(x_n) = b$ for any sequence $x_n \to x_0$ is very essential here since it may happen so

that for one sequence of values of the variable this limit exists and is equal to b and for some other sequence the limit does not exist, or exists but is different from b.

The definition of the limit of a function given above is equivalent to the following: the number b is the limit of the function $y = f(x)$ as $x \to x_0$ if, whatever the neighbourhood $(b - \alpha, b + \alpha)$ of the number b, all values of the function belong to this neighbourhood for the values of x sufficiently close to x_0.

We shall not give here the proof of the equivalence of these two definitions of the limit of a function.

Example 1. Calculate the limit of the function $y = \frac{3x-4}{x^2+2}$ as $x \to 2$ using the definition of the limit of a function.

Solution. Assume that x_n is a sequence of values of the variable x which converges to the number 2. Then the nth term of the corresponding sequence of values of the function has the form $y_n = \frac{3x_n-4}{x_n^2+2}$. Let us calculate the limit of the sequence y_n as $x_n \to 2$:

$$\lim_{x_n \to 2} (3x_n - 4) = 3 \cdot 2 - 4 = 2,$$

$$\lim_{x_n \to 2} (x_n^2 + 2) = 2^2 + 2 = 6,$$

$$\lim_{x_n \to 2} \frac{3x_n-4}{x_n^2+2} = \frac{-2}{6} = \frac{1}{3}.$$

The result obtained is valid for any sequence x_n converging to the number 2, and, consequently, according to the definition, $\lim\limits_{x \to 2} \frac{3x-4}{x^2+2} = -\frac{1}{3}$.

Example 2. Calculate $\lim\limits_{x \to 2} \frac{x^2-4}{x-2}$.

Solution. For $x = 2$ the function is meaningless. Here is a preliminary remark: the concept of the limit of a function as $x \to x_0$ is connected with the behaviour of the function in the vicinity of the point x_0, except for the point x_0 itself (see the definition). The fraction $\frac{x^2-4}{x-2}$ is identically equal to the expression $x + 2$ at all points except for the point $x = 2$. Consequently, in the vicinity of the point $x = 2$ as well the

functions $\frac{x^2-4}{x-2}$ and $x+2$ coincide and

$$\lim_{x\to 2}\frac{x^2-4}{x-2}=\lim_{x\to 2}(x+2)=4.$$

Thus, although the function $\frac{x^2-4}{x-2}$ is meaningless at the point $x=2$, this does not exclude the possibility of the existence of the limit of the function as $x\to 2$.

Example 3. Since the calculation of the limit of the function $f(x)$ as $x\to x_0$ reduces to the calculation of the limit of the sequence $f(x_n)$ as $x_n\to x_0$, it is easy to formulate and prove the theorems on the limit of the sum, the product and the quotient of two functions which have limits as $x\to x_0$.

By way of example, we shall formulate and prove the theorem on the limit of the sum of two functions:

If $\lim\limits_{x\to x_0} f(x)=a$ *and* $\lim\limits_{x\to x_0} g(x)=b$, *then* $\lim\limits_{x\to x_0}(f(x)+g(x))=a+b$.

Proof. Let us consider a sequence x_n converging to x_0. According to the hypothesis, the sequences $f(x_n)\to a$ and $g(x_n)\to b$. We apply Theorem 5 on the limit of the sum of sequences to the sequence $f(x_n)+g(x_n)$. We obtain $\lim\limits_{x_n\to x_0}(f(x_n)+g(x_n))=a+b$. Since our arguments are valid for any sequence x_n which converges to x_0, it follows that $\lim\limits_{x\to x_0}(f(x)+g(x))=a+b$.

Formulate and prove similar theorems on the limit of the product and the quotient of two functions which have a limit as $x\to x_0$.

Exercise

Proceeding from the theorem on the sum, the product and the quotient of two functions, calculate the following limits:

1. $\lim\limits_{x\to 1}(3x^2-2x+7)$. 2. $\lim\limits_{x\to -1}\frac{x^2-3x+4}{x^2-5}$.

1.7. Continuity of Functions

We can introduce the concept of continuity of a function proceeding from graphic representations, namely, a function is *continuous* if its graph consists of one continuous curve.

The concept of the limit of a function just introduced makes it possible to give a more exact definition of continuity.

Let the function $f(x)$ be defined on an interval which contains a point x_0 in its interior or on its boundary. The function $f(x)$ is *continuous at the point* x_0 if there is a limit $\lim_{x \to x_0} f(x)$ and this limit is equal to $f(x_0)$. In terms of geometry, the continuity of the function at the point x_0 means that its graph does not suffer a discontinuity at the point x_0.

A function is *continuous on an interval* if it is continuous at all points of that interval.

A point at which a function is not continuous but is defined in its neighbourhood is a *point of discontinuity*.

Points at which a function is not defined (but is defined in the neighbourhood) are also points of discontinuity. Sometimes at these points there is an *inessential* or *removable discontinuity* which is eliminated when the definition of the function at this point is completed as needed.

Thus the function $\frac{x^2-1}{x-1}$ is not defined at the point $x = 1$. However, for $x \neq 1$ it is equal to $x + 1$ and as $x \to 1$ it has a limit equal to 2. "Completing" the definition of this function, i.e. assuming the number 2 to be its value at $x = 1$, we get a continuous function. It is equal to $x + 1$ for all values of x including 1.

The function sgn x is not defined for $x = 0$ either and has a nonremovable discontinuity at this point. As $x \to 0$ from the right (i.e. from the side of positive values), sgn $x = 1$ and the limit is equal to 1, as $x \to 0$ from the left, the limit of the function sgn x is -1 and so, as $x \to 0$, the limit of sgn x does not exist and it is impossible to "complete" the definition of this function for $x = 0$ so that it would become continuous.

◀ The function $\frac{2^{1/x}+2}{2^{1/x}+1}$ suffers a similar discontinuity at the point $x = 0$ at which the function is not defined. Indeed, as $x \to 0$ from the left, $1/x \to -\infty$ so that $2^{1/x} \to 0$ and $\frac{2^{1/x}+2}{2^{1/x}+1}$ has a limit equal to 2. As $x \to 0$ from the right, $1/x \to +\infty$, $2^{1/x} \to \infty$, $2^{-1/x} \to 0$

and then

$$\frac{2^{1/x}+2}{2^{1/x}+1}=\frac{2^{1/x}(1+2\cdot 2^{-1/x})}{2^{1/x}(1+2^{-1/x})}=\frac{1+2\cdot 2^{-1/x}}{1+2^{-1/x}}\to 1.$$

Thus the function in question has a limit 2 as $x\to 0$ from the left and a limit 1 as $x\to 0$ from the right. These limits are different and so the discontinuity here cannot be removed. ▶

The function $\frac{1}{x-1}$ has another kind of discontinuity at the point $x=1$. It is clear that $\frac{1}{x-1}\to-\infty$ if x tends to 1 from the left and $\frac{1}{x-1}\to+\infty$ if $x\to 1$ from the right. A discontinuity of this kind is typical for fractional rational functions at points where the denominator vanishes. The function may also tend to discontinuity of one sign. The discontinuity of the function $\frac{1}{(x-1)^2}$ at the point $x=1$ can serve as an example of this statement.

Now we shall explain the concept of continuity of a function in other terms. The relation $x\to x_0$ is equivalent to the fact that $x-x_0$ is an infinitesimal variable. Respectively, the relation $f(x)\to f(x_0)$ is equivalent to the fact that $f(x)-f(x_0)$ is an infinitesimal. The differences $x-x_0$ and $f(x)-f(x_0)$ are *increments* of the independent variable x and of the function $f(x)$ upon a passage from x_0 to x. Thus we can formulate the definition of continuity of the function $f(x)$ at the point x_0 as follows: the function $f(x)$ is continuous at the point x_0 if an infinitesimal increment $x-x_0$ of the independent variable is associated with an infinitesimal increment $f(x)-f(x_0)$ of the function. We can prove that integral rational functions, fractional rational functions as well as power, exponential and logarithmic functions are continuous throughout the domain of definition.

Consider the following examples.

Example 1. If the function $f(x)$ is known to be continuous at the point x_0, then the problem of calculating the limit of the function $f(x)$ as $x\to x_0$ is trivial since according to the definition of the continuity of the function $f(x)$ at the point x_0

$$\lim_{x\to x_0} f(x)=f(x_0),$$

i.e. the calculation of the limit reduces to the calculation of the value of the function at the point x_0.

Assume, for instance, that we have to find $\lim\limits_{x\to 2} \frac{x^2-3}{x-1}$. As was stated above (without proof), all fractional rational functions are continuous at every point of the domain of definition. The value $x = 2$ belongs to the domain of definition of the function $\frac{x^2-3}{x-1}$. Consequently, the required limit is equal to the value of the function at the point $x = 2$, i.e.

$$\lim_{x\to 2} \frac{x^2-3}{x-1} = \frac{2^2-3}{2-1} = 1.$$

Example 2. Find $\lim\limits_{x\to 3} \frac{x-2-\sqrt{x-2}}{x^2-9}$.

Solution. The function $\frac{x-2-\sqrt{x-2}}{x^2-9}$ is meaningless at the point $x=3$ and therefore cannot evidently be considered to be continuous at the point $x=3$. However, we know that this does not exclude the possibility of the existence of a limit as $x \to 3$. We transform the expression under the limit sign:

$$\frac{x-2-\sqrt{x-2}}{x^2-9} = \frac{(x-2-\sqrt{x-2})(x-2+\sqrt{x-2})}{(x^2-9)(x-2+\sqrt{x-2})}$$

$$= \frac{x^2-4x+4-x+2}{(x^2-9)(x-2+\sqrt{x-2})} = \frac{x^2-5x+6}{(x^2-9)(x-2+\sqrt{x+2})}$$

$$= \frac{(x-3)(x-2)}{(x-3)(x+3)(x-2+\sqrt{x-2})} = \frac{x-2}{(x+3)(x-2+\sqrt{x-2})}.$$

Comparing the initial expression with that obtained, we see that a neighbourhood of the point $x = 3$ can be found at whose every point, except for the point $x = 3$ itself, the values of these expressions coincide. This means that as $x \to 3$, the limits of the functions defined by these expressions are equal but the expression obtained is meaningful at the point $x = 3$ and the corresponding function is continuous at the point $x = 3$ so that, as $x \to 3$, its limit is equal to its

value at this point. The final result is

$$\lim_{x\to 3}\frac{x-2-\sqrt{x-2}}{x^2-9}=\lim_{x\to 3}\frac{x-2}{(x+3)(x-2+\sqrt{x-2})}$$
$$=\frac{3-2}{(3+3)(3-2+\sqrt{3-2})}=\frac{1}{12}.$$

A situation similar to that described in Example 2 often arises in calculations of derivatives.

Example 3. Let us consider the graphs of functions which have discontinuities of various kinds at some points.

(a) $y = \operatorname{sgn} x$. See Fig. 1.2. The arrows at the ends of the rectilinear portions of the graph mean that for $x = 0$ the function is not defined but, approaching $x = 0$, it tends to the value $y=1$ from the right and to the value $y = -1$ from the left.

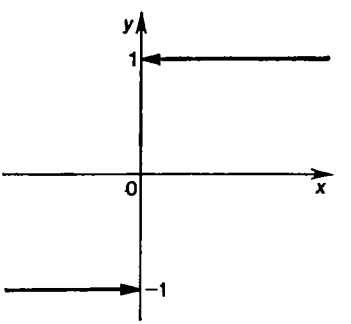


Fig. 1.2

◀ (b) $y=\dfrac{2^{1/x}+2}{2^{1/x}+1}$ or, else, $y=1+\dfrac{1}{2^{1/x}+1}$. See Fig. 1.3. At the point $x=0$ the function is not defined and has a discontinuity of the same kind as in item (a). As x increases indefinitely ($x\to\infty$), the values of the function approach 1.5 from below and, as $x\to -\infty$, the values of the function approach 1.5 from above. ▶

(c) $y=\begin{cases} x & \text{for } x\leqslant 1,\\ x+1 & \text{for } x>1.\end{cases}$ See Fig. 1.4. The function is defined at the point $x=1$ but suffers a discontinuity of the same kind as those in examples 1 and 2.

(d) $y=\dfrac{x^2-1}{x-1}$. See Fig. 1.5. The function is meaningless at the point $x=1$ but $\lim\limits_{x\to 1}\dfrac{x^2-1}{x-1}=2$. At the point $x=1$ the function has a removable discontinuity.

(e) $y=\dfrac{1}{x^2}$. See Fig. 1.6. The function is meaningless at the point $x = 0$. At this point the function has a discontinuity typical of fractional rational functions.

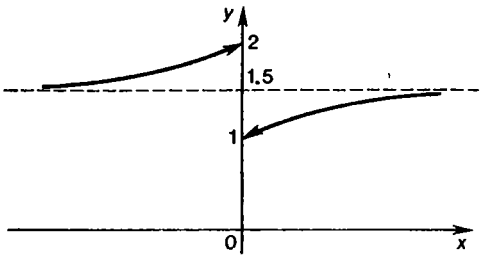


Fig. 1.3

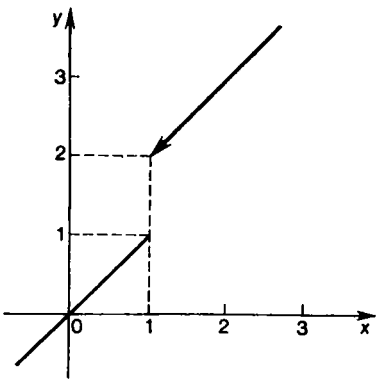


Fig. 1.4

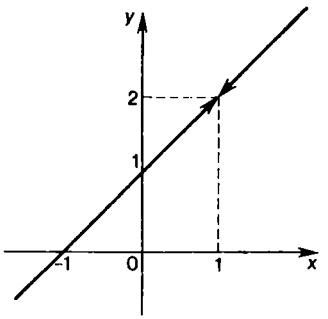


Fig. 1.5

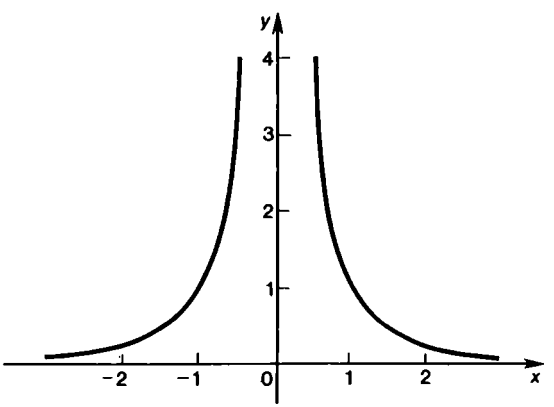


Fig. 1.6

1.8. A More Exact Definition of the Concept of a Derivative

The linear function $l(x) = kx + b$ possesses the following property: an increment of a linear function is **proportional** to an increment of an independent variable and the proportionality factor is equal to the slope k. Indeed, $l(x) - l(x_0) = kx + b - (kx_0 + b) = k(x - x_0)$.

According to the main principle stated in 1.1, a "good" function $f(x)$ is almost linear on a small interval, say, on the interval adjoining the point x_0, and the smaller the interval, the closer it is to a linear function. Assume that k is the slope of this linear function, i.e. the value $f'(x_0)$ of the derivative of the function at the point x_0. Then the increment of the function $f(x)$ is **almost proportional** to the increment of the independent variable with the proportionality factor k, i.e. we have an approximate equality

$$\frac{f(x) - f(x_0)}{x - x_0} \cong k$$

and this equality is the more exact the smaller $x - x_0$. This means that

$$f'(x_0) = k = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

The increment $x - x_0$ of the independent variable is usually designated as Δx (this designation is not very apt since it does not include the initial point x_0). Correspondingly, the increment $f(x) - f(x_0)$ of the function $y = f(x)$ is designated as Δy. In these designations

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x},$$

or, in other words, *the value of the derivative $f'(x_0)$ of the function $f(x)$ at the point x_0 is equal to the limit of the ratio of the increment of the function to the corresponding increment of the independent variable when the latter increment tends to zero.* This formulation implies that the value x_0 from which the increment of the independent variable is reckoned remains constant.

This formulation is, in essence, the definition of the value of the derivative at a given point which refines the geometric definition given in 1.1.

When we make calculations, it is convenient to transform the formula $f'(x_0) = \lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$. It is clear that $x = x_0 + \Delta x$ and $\Delta y = f(x) - f(x_0) = f(x_0 + \Delta x) - f(x_0)$. Therefore

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}.$$

If, in accordance with what was said in 1.1, we consider the derivative $f'(x)$ as a function of the same argument x, we get the following expression for this function:

$$f'(x) = \lim_{\Delta x \to 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}.$$

It should be remembered that in the last formula x must be regarded as a constant and make a passage to the limit as the variable Δx tends to zero.

If the function $f(x)$ is denoted by one letter, say, y, then the derivative is designated as y', to match the function.

Theorem 8. *If the derivative $f'(x_0)$ exists at some point x_0, then the function $f(x)$ is continuous at this point.*

Proof. Indeed, if $\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$ has a limit $f'(x_0)$ as $\Delta x \to 0$, then $\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} = f'(x_0) + \alpha$, where $\alpha \to 0$ as $\Delta x \to 0$. Consequently, the quantity $\Delta y = f(x_0 + \Delta x) - f(x_0) = f'(x_0)\,\Delta x + \alpha\,\Delta x$ also tends to zero as $\Delta x \to 0$ since both terms on the right-hand side tend to zero. But the fact that $\Delta y \to 0$ as $\Delta x \to 0$ exactly means that $f(x)$ is continuous at the point x_0.

◀ Generally speaking, the converse statement is not true. Not every function which is continuous at the point x_0 has a derivative at this point. For instance, we can easily verify that the function $y = |x|$ does not have a derivative at the point $x = 0$. And what is more, we can construct a function continuous on an interval which would not have a derivative at any point of the interval. It can be visualized as follows: the graph of the function is a wavy line. Each wave consists of smaller waves which, in turn, consist of still smaller waves and so on. Not at any point will the part of the graph in the neighbourhood of that point become linear and the derivative will not exist at any point. It is not

simple to formalize the description of such a function which contains a limiting process replacing the words "and so on".

We can now give a more exact definition of a smooth function, namely, the function $f(x)$ is *smooth* on an interval if the derivative $f'(x)$ exists at all points of that interval and is continuous. The continuity of the derivative which is the slope of the tangent means that the tangent to the graph changes its direction continuously upon a continuous variation of the abscissa of the point of tangency. This is in an absolute accord with the graphic representation of a continuous smooth curve without inflections. ▶

We shall see in the next chapter that integral and fractional rational functions as well as exponential and logarithmic functions have continuous derivatives in the whole natural domain of existence of each of these functions.

Here are some examples of calculation of derivatives.

Example 1. Find the derivative of the function $y = x^2$.

Solution.

$$y' = \lim_{\Delta x \to 0} \frac{(x+\Delta x)^2 - x^2}{\Delta x} = \lim_{\Delta x \to 0} \frac{x^2 + 2x\,\Delta x + (\Delta x)^2 - x^2}{\Delta x}$$
$$= \lim_{\Delta x \to 0} (2x + \Delta x) = 2x.$$

Example 2. Find the derivative of the function $y = \frac{1}{x}$.

Solution.

$$y' = \lim_{\Delta x \to 0} \frac{\frac{1}{x+\Delta x} - \frac{1}{x}}{\Delta x} = \lim_{\Delta x \to 0} \frac{x - (x+\Delta x)}{\Delta x \cdot x\,(x+\Delta x)}$$
$$= \lim_{\Delta x \to 0} \left[-\frac{1}{x\,(x+\Delta x)} \right] = -\frac{1}{x^2}.$$

Example 3. Find the derivative of the function $y = \sqrt{x}$, $x > 0$.

Solution.

$$y' = \lim_{\Delta x \to 0} \frac{\sqrt{x+\Delta x} - \sqrt{x}}{\Delta x}$$
$$= \lim_{\Delta x \to 0} \frac{(\sqrt{x+\Delta x} - \sqrt{x}) \cdot (\sqrt{x+\Delta x} + \sqrt{x})}{(\sqrt{x+\Delta x} + \sqrt{x})\,\Delta x}$$
$$= \lim_{\Delta x \to 0} \frac{(x+\Delta x) - x}{(\sqrt{x+\Delta x} + \sqrt{x})\Delta x} = \lim_{\Delta x \to 0} \frac{1}{\sqrt{x+\Delta x} + \sqrt{x}} = \frac{1}{2\sqrt{x}}.$$

Note that for $x = 0$ the derivative does not exist.

1.9. The Equation of a Tangent to the Graph of a Function

Consider a function $f(x)$ which has a derivative at a point x_0. We know that $f'(x_0)$ is the slope of the tangent to the graph at the point with abscissa x_0. Therefore the equation of the tangent line has the form $y = f'(x_0)\,x + b$. The point of tangency has an abscissa x_0 and ordinate $f(x_0)$. These coordinates must satisfy the equation of the tangent line whence follows an equality $f(x_0) = f'(x_0)\,x_0 + b$, from which we get $b = f(x_0) - f'(x_0)\,x_0$. Substituting this value into the equation of the tangent line, we obtain it in the form

$$y = f'(x_0)\,x + f(x_0) - f'(x_0)x_0,$$

or, after obvious transformations, in the form

$$y - f(x_0) = f'(x_0)\,(x - x_0).$$

Example 1. Write the equation of a tangent to the graph of the function $y = \sqrt{x}$ at the point with abscissa $x = 4$.

Solution. To write the equation of a tangent to the graph

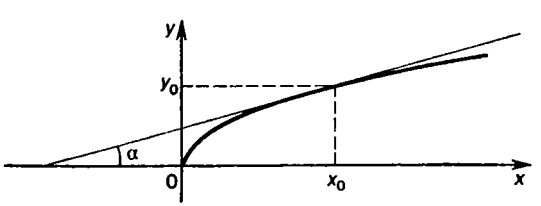


Fig. 1.7

of the function $f(x)$, we have to find three numbers: x_0, the abscissa of the point of tangency, $f(x_0)$, the ordinate of the point of tangency and $f'(x_0)$, the slope of the tangent ($f'(x_0) = \tan\alpha$, see Fig. 1.7). By the hypothesis $x_0 = 4$, $f(x_0) = \sqrt{4} = 2$, $f'(x) = \frac{1}{2\sqrt{x}}$ and $f'(4) = \frac{1}{4}$. Substituting the numbers found into the equation of the tangent line, we get $y - 2 = \frac{1}{4}(x - 4)$, or $y = \frac{1}{4}x + 1$.

Example 2. Using the result of the preceding problem, find the approximate value of $\sqrt{4.03}$.

Solution. In accordance with the main principle of differential calculus, on a small interval of variation of the argument the graph of a smooth function almost coincides with a portion of a straight line and the tangent at the point with abscissa x_0 is the straight line which is the closest to the graph of the function in the vicinity of this point. Therefore, using the equation of the tangent to the graph of the function $y = \sqrt{x}$ at the point with abscissa $x_0 = 4$, we can find an approximate value of this function in the vicinity of the point of tangency:

$$\sqrt{4.03} \cong \frac{1}{4} \cdot 4.03 + 1 = 1.0075 + 1 = 2.0075.$$

For comparison: the root $\sqrt{4.03}$ is equal to 2.007486 with an accuracy to within 0.000001.

In this example we speak only of the theoretical possibility of obtaining an approximate value of the function from the equation of the tangent line. In the next chapters we shall present the arguments which will allow us to estimate the error of approximation.

1.10. The Rate of Variation of a Function

Consider, for example, a function $y = x^2$ and its graph for $x \geqslant 0$ (Fig. 1.8). When we look at the graph, we see that at the beginning, about 0, the function increases slowly, and then the increase becomes faster. Let us refine this idea of the rate of variation of the function.

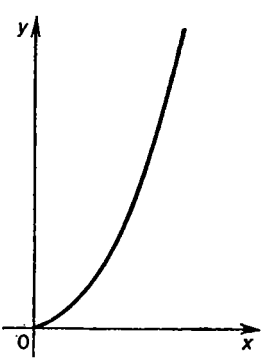


Fig. 1.8

It is natural to consider the fraction $\frac{f(x+\Delta x)-f(x)}{\Delta x}$, i.e. the ratio of the variation of the function to that of the argument, to be the average rate of variation of the function $f(x)$ as the independent variable changes from x to $x+\Delta x$. If, as $\Delta x \to 0$, the average rate has a limit, then this limit is assumed to be the true rate of variation of the function at the point x. But this limit is nothing but the derivative $f'(x)$ of the function $f(x)$.

We have thus obtained the following significant interpretation of a derivative: *the derivative of the function $f(x)$ at a point is the rate of variation of the function at that point.*

This result is in full accord with the idea of the derivative as the slope of the tangent, i.e. it is clear that the rate of the increase of a function is faster where the ascent of the graph of the function is steeper, i.e. where the slope of the tangent is larger. Respectively, the rate of the decrease of a function is faster where the descent of the graph of the function is steeper, i.e. where the negative, in this case, slope of the tangent is greater in absolute value.

Exercises

1. Looking at the graph of the function $y = f(x)$ (Fig. 1.9), say, at which of the points x_1, x_2, x_3 the derivative of the function $f(x)$ is larger.

2. Looking at the graph of the function $y = g(x)$ (Fig. 1.10), say, at which of the points x_1, x_2, x_3 the derivative of the function $g(x)$

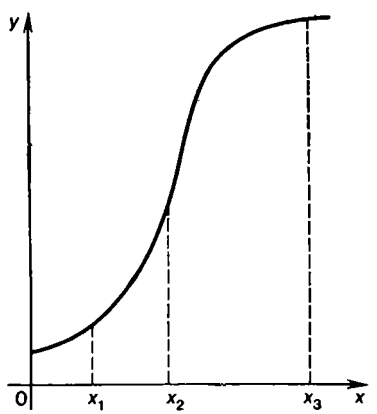


Fig. 1.9

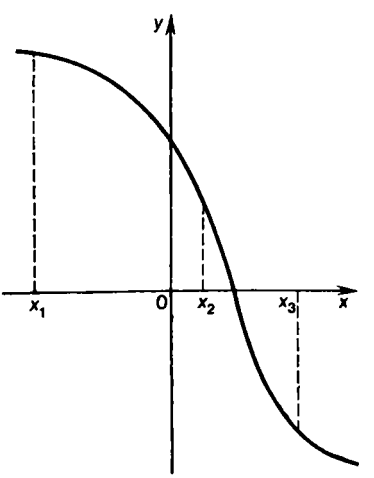


Fig. 1.10

is smaller; at which of these points the derivative is larger in absolute value.

3. Formulate the answers to the two preceding problems in the terms "the rate of increase" and "the rate of decrease".

1.11. The Velocity of Mechanical Motion of a Particle in a Straight Line

Assume that a particle moves in a straight line so that it changes its position with time. We regard this straight line as the number axis, i.e. choose a scale, a reference point and a positive direction. Then the position of the particle is defined by its coordinate and this coordinate changes with time thus being a function of time. Assume that y is the coordinate of the moving particle, t is time (in seconds or in some other time units convenient for a particular case). The *equation of motion* is the notation $y = f(t)$ which shows how the coordinate changes with time (the function $f(t)$ is assumed to be given).

The motion is *uniform* if the coordinate of the particle changes in proportion to the time of motion. The proportionality factor, i.e. the path traversed in unit time (with the plus or minus sign depending on whether the motion is in the positive or negative direction), is the *velocity of uniform motion*. If v is the velocity and y_0 is the coordinate of the particle at the initial moment, then, evidently, $y - y_0 = vt$ so that the equation of motion is

$$y = y_0 + vt.$$

Thus, when the motion is uniform, the coordinate of the moving particle is a linear function of time.

However, in nature and in engineering we often have to deal with nonuniform motion. The qualitative aspect of this motion is usually clear, i.e. we see when the nonuniform motion of a particle is faster and when it is slower, but to understand the quantitative aspect of the problem, it is necessary to direct our attention to the concepts of differential calculus as we did when we tried to understand the concept of the rate of variation of a function. We shall repeat our arguments in application to the situation in question.

The *average velocity of motion* on the given interval of time is the quotient of the division of the rate of variation of the coordinate of the moving particle (if the particle moves in the positive direction, this value coincides with the length of the path traversed) by the value of the time interval. Let $y = f(t)$ be the equation of motion. We shall con-

sider the time interval from the moment t to the moment $t + \Delta t$. The variation of the coordinate is evidently

$$\Delta y = f(t + \Delta t) - f(t).$$

The length of the time interval is Δt. The average velocity on the interval being considered is

$$\frac{\Delta y}{\Delta t} = \frac{f(t+\Delta t) - f(t)}{\Delta t}.$$

When the length Δt of the time interval decreases indefinitely, the average velocity tends to the limit (if the motion is smooth), which is exactly the *velocity of motion at time t*. But this limit is a derivative $f'(t)$ of the function $f(t)$ which defines the equation of motion.

Thus the *velocity of motion with the equation* $y = f(t)$ *at the moment t is equal to the value of the derivative* $f'(t)$ *at this time moment.*

The great mathematician, physicist and astronomer Isaac Newton (1643-1727), one of the founders of differential calculus, based his calculus on the concept of the velocity of motion.

The velocity of nonuniform motion varies with time. The rate of variation of velocity is called *acceleration*. The acceleration is equal to the derivative of the velocity as a function of time. The concept of acceleration is of special importance in physics and mechanics by virtue of Newton's second law. The gist of this law is that when a particle moves under the action of force, the acceleration is proportional to the motive force applied and inversely proportional to the mass of the particle.

The motion defined by the equation

$$y = at^2 + bt + c$$

is one of the most important examples of motion. Let us find the velocity and acceleration for this motion. We calculate the velocity:

$$v = \lim_{\Delta t \to 0} \frac{\Delta y}{\Delta t} = \lim_{\Delta t \to 0} \frac{a(t+\Delta t)^2 + b(t+\Delta t) + c - (at^2 + bt + c)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{at^2 + 2at\,\Delta t + a(\Delta t)^2 + bt + b\,\Delta t + c - at^2 - bt - c}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{2at\,\Delta t + a\,(\Delta t)^2 + b\,\Delta t}{\Delta t}$$

$$= \lim_{\Delta t \to 0} (2at + a\,\Delta t + b) = 2at + b.$$

Thus $v = 2at + b$.

We shall now calculate the acceleration designating it as j:

$$j = v' = \lim_{\Delta t \to 0} \frac{\Delta v}{\Delta t} = \lim_{\Delta t \to 0} \frac{2a\,(t + \Delta t) + b - (2at + b)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{2a\,\Delta t}{\Delta t} = 2a.$$

Thus, when motion is defined by the equation $y = at^2 + bt + c$, the acceleration does not depend on time and is equal to $2a$. The motion with constant acceleration is a *uniformly accelerated motion*. Thus the motion defined by the equation $y = at^2 + bt + c$ is a uniformly accelerated motion. The coefficient a is equal to half the acceleration.

It is easy to interpret the other coefficients. The coordinate y_0 of the particle and its velocity v_0 at the moment $t = 0$ are the *initial coordinate* and *initial velocity*. It is clear that $y_0 = a \cdot 0^2 + b \cdot 0 + c = c$ and $v_0 = 2a \cdot 0 + b = b$. Thus the coefficients b and c are equal to the initial velocity and the initial coordinate respectively.

Thus the equation $y = at^2 + bt + c$ is the equation of uniformly accelerated motion with acceleration $2a$ for the initial coordinate c and initial velocity b. We shall prove later on that any uniformly accelerated motion has an equation of the form $y = at^2 + bt + c$.

Example 1. Figure 1.11 shows the graph of motion of the particle along a coordinate line, namely, the graph of variation of the coordinate of the particle as a function of time. The units of measurement and the scale units are time along the abscissa axis, with the division value of 1 s, and the coordinate along the axis of ordinates, with the division value of 1 m.

Using that graph, determine

(1) the initial position of the particle,

(2) the approximate value of the initial velocity,

(3) intervals of time in which the particle moved practically uniformly,

(4) time intervals in which the particle moved with deceleration,

(5) the "stop" moment, i.e. the moment at which the velocity was equal to zero,

(6) time intervals in which the particle moved with acceleration, i.e. the absolute value of the velocity increased,

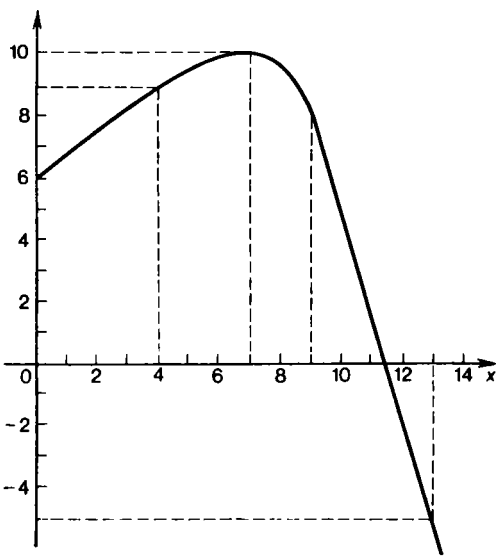


Fig. 1.11

(7) moments of time in which the absolute value of the velocity was the greatest,

(8) the path that the particle traversed from the beginning of motion to the moment $t = 13$ s.

Solution. The initial position of the particle can be found directly from the figure, $y(0) = 6$.

We shall first answer the third question. Applying a ruler, we can make sure that the portions of the graph on the intervals from 0 to 4 and from 9 to 13 practically coincide with portions of straight lines. This means that in these time intervals the particle moved practically uniformly, i.e. with constant velocity. In the interval from 0 to 4 this velocity is equal to the initial velocity and can be found by dividing

the increment of the coordinate by the length of the interval of time $v(0) \cong \frac{9-6}{4} = 0.75$ m/s. We note that on the interval from 4 to 7 the graph becomes less steep. This means that the absolute value of the velocity decreases in this interval, i.e. the particle moves with deceleration. In the interval from 7 to 9 the graph becomes steeper again but the graph descends. This means that the velocity has changed sign, but the absolute value of the velocity increases, i.e. the particle moves with acceleration. It can be seen by eye that on the interval from 9 to 13 the graph is steeper than on the interval from 0 to 4. This means that on the interval from 9 to 13 the velocity is greater in absolute value than on the interval from 0 to 4. At the point with abscissa 7 the tangent to the graph is parallel to the abscissa axis and this means that the derivative of the ordinate as a function of time is zero at this moment, i.e. at the moment $t = 7$ s the instantaneous velocity of motion is zero, this is the moment of the stop. The path the particle traversed during the time from the start of motion to $t = 13$ s can be calculated directly from the graph. Moving in the positive direction, the particle traversed a path $S_1 = 10-6 = 4$ m during the time from $t = 0$ to $t = 7$. Moving in the negative direction, the patricle traversed a path $S_2 = 10 - (-5) = 15$ m during the time from $t = 7$ to $t = 13$. The whole path traversed by the particle during the time from $t = 0$ to $t = 13$ is $S = S_1 + S_2 = 4 + 15 = 19$ m.

Note that the path traversed by the particle during the time from $t = 0$ to $t = 13$ must not be confused with the change (increment) of the coordinate of the particle during this time $\Delta y = y(13) - y(6) = +5 - 6 = -11$ m.

1.12. The Differential of a Function

Alongside the derivative, the differential of a function is one of the most important concepts in mathematics.

The *differential* dy of the function $y = f(x)$ is the product of the derivative $f'(x)$ by the increment Δx of the independent variable:

$$dy = f'(x)\,\Delta x.$$

It is evident that $y' = 1$ for $y = x$. Indeed, here

$$\Delta y = \Delta x, \quad \frac{\Delta y}{\Delta x} = 1 \text{ and } y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 1.$$

Therefore $dx = 1 \cdot \Delta x = \Delta x$ so that for the independent variable the increment and the differential coincide. Taking this fact into account, we can write

$$dy = f'(x)\, dx$$

for $y = f(x)$.

In the formal definition of the differential no restrictions are imposed on dx. However, in the applications of this concept to various problems of mathematics, physics and, in general, natural sciences the quantity dx *is considered to be infinitesimal* (and thus variable). Also take note of the fact that the derivative is equal to the ratio of the differentials: $y' = \frac{dy}{dx}$. This fraction is often used to designate a derivative.

Let us study the differential more closely, trying to understand its essence and visualize it. Recall that the derivative $y' = f'(x)$ is the limit of the ratio $\frac{\Delta y}{\Delta x}$ of the increments as $\Delta x \to 0$. Therefore the ratio of the increments itself differs from its limit $f'(x)$ by an infinitesimal which we shall designate as α. Thus

$$\frac{\Delta y}{\Delta x} = f'(x) + \alpha, \text{ where } \alpha \to 0 \text{ as } \Delta x \to 0.$$

Multiplying this equality by Δx, we obtain

$$\Delta y = f'(x)\, \Delta x + \alpha \Delta x = dy + \alpha dx.$$

The second term αdx on the right-hand side of this relation *tends to zero appreciably faster* than $dx = \Delta x$. Thus dy is a part of the increment Δy and differs from it by a term which tends to zero appreciably faster than Δx. Therefore the differential dy is called the *principal part* of the increment Δy. Furthermore, for a fixed x the differential dy is proportional to the increment dx of the independent variable. All we have said allows us to formulate the following statement: *the differential dy of the function $y = f(x)$ is the principal part of the increment Δy proportional to Δx.*

The converse is also true, namely, if $k\,dx$ is the principal part of the increment Δy of the function $y = f(x)$, which is proportional to dx, then it is the differential dy. Indeed, let $k\,dx = k\,\Delta x$ be this principal part, i.e. $\Delta y = k\,\Delta x + \lambda$, where λ is an infinitesimal tending to zero appreciably faster than Δx. Then $\frac{\Delta y}{\Delta x} = k + \frac{\lambda}{\Delta x}$. But $\frac{\lambda}{\Delta x} \to 0$ as $\Delta x \to 0$ since λ tends to zero appreciably faster than Δx. Therefore, $\lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = k$, i.e. $k = y'$, and

$$k\Delta x = y'dx = dy$$

indeed.

Thus the property of a differential to be the principal part of an increment, which is proportional to the increment of the independent variable, fully characterizes the differential and can be accepted as the definition of the differential.

The definition of a differential as the principal linear part of an increment makes it possible, in many cases, to calculate the differential of a function without calculating the derivative. For instance, let $y = x^3$. Then

$$\begin{aligned}\Delta y &= (x+\Delta x)^3 - x^3 = x^3 + 3x^2\Delta x + 3x(\Delta x)^2 + (\Delta x)^3 - x^3 \\ &= 3x^2\Delta x + 3x\,(\Delta x)^2 + (\Delta x)^3.\end{aligned}$$

The second $3x\,(\Delta x)^2$ and the third $(\Delta x)^3$ terms tend to zero appreciably faster than Δx and the first term is proportional to Δx. Therefore the first term is precisely the required differential: $dy = 3x^2\Delta x = 3x^2dx$.

The isolation of a differential from the increment of a function as the principal part defines the main application of this concept, namely, it often allows us to find the differential of a function even when the function itself is yet unknown.

Let us consider an example of this kind which is significant for what follows (we shall repeatedly return to it). Let $y = f(x)$ be a positive function defined on the interval $[a, b]$. Its graph lies above the abscissa axis. We designate as $F(x)$ the area of a figure (Fig. 1.12) bounded by the graph of the function $f(x)$, the abscissa axis and vertical straight lines with abscissas a and x. The area $F(x)$ is indeed a function of the abscissa x of the right-hand wall. We assign an increment dx to the number x. Then the area $F(x)$ gets an incre-

ment which is the area of a strip dx wide. This strip is close to a rectangle with an altitude $y = f(x)$. It is clear "by eye" that the area of the strip differs from the area $y\,dx$ of the

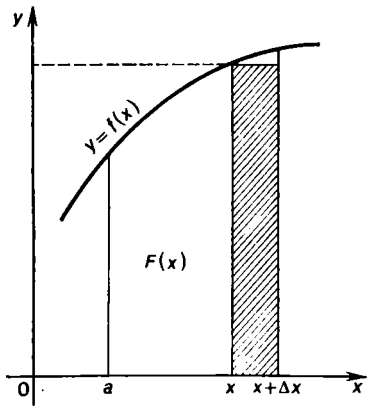


Fig. 1.12

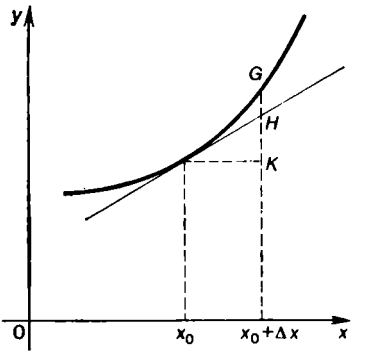


Fig. 1.13

rectangle by the quantity λ which tends to zero appreciably faster than dx. This assertion can be rigorously proved. Thus

$$\Delta F(x) = y\,dx + \lambda.$$

Thus $y\,dx$ is the principal part of the increment $\Delta F(x)$ and is proportional to dx. Consequently, $d(F(x)) = y\,dx$.

We have shown by this example the possibility of calculating a differential (and, consequently, a derivative) of a function which is yet unknown. To find it, we must perform an operation converse to the calculation of a differential, namely, the operation of finding a function from a given differential. The operation of finding a differential (and a derivative) of a given function is known as *differentiation*. The converse operation, that of finding a function from a given differential, is known as *integration*. Thus, to find the area $F(x)$ we must integrate the differential $y\,dx$.

The differential has a simple geometrical interpretation. Figure 1.13 showsthe graph of the function $y = f(x)$ and a tangent to it at a point with abscissa x_0. Let us compare the increment of the function $f(x)$ with that of the ordinate

of the point on the tangent. The increment Δy of the function is represented by the length of the segment GK. The increment of the ordinate of the point on the tangent is equal to the length of the segment HK. Proceeding from the equation of the tangent $y - y_0 = f'(x_0)(x - x_0)$, we find that the increment of the ordinate along the tangent is

$$y - y_0 = f'(x_0)(x - x_0) = f'(x_0)\,\Delta x = dy.$$

Thus the length of the segment HK is equal to the value of the differential dy at the point x_0. The difference $\Delta y - dy$ is equal to the length of the segment GH. We know that this difference is an infinitesimal which tends to zero appreciably faster than Δx. This can be easily seen "by eye".

Let us now recall the "main principle": a segment of a smooth curve looks like a segment of a straight line (a tangent) and is the closer to the straight line, the smaller the segment, for a small variation of the argument a smooth function is close to a linear function and the closer to it, the smaller the interval of variation of the argument. It is clear from what we have discussed that the differential of a function at a given point is equal to its increment with an accuracy to within the replacement of the function by a linear function which is the closest to it in the neighbourhood of this point.

1.13. The Differential of a Function of a Function

Let $y = F(u)$ be a function of the variable u, which, in turn, is a function $u = f(x)$ of the independent variable x. Thus $y = F(f(x))$. The function y thus defined is a *function of a function*, or a *composite function*.

Let us find the differential of the function y. Assume that the independent variable x has got an increment Δx. We denote the corresponding increments of the variables u and y by Δu and Δy. We know that if the variable u has got an increment Δu, then the function $y = F(u)$ will get an increment

$$\Delta y = F'(u)\,\Delta u + \alpha \Delta u,$$

where α is a variable tending to zero together with Δu. Furthermore,

$$\Delta u = f'(x)\,\Delta x + \beta\,\Delta x = du + \beta\,\Delta x,$$

where β is a variable tending to zero together with Δx. Note that as soon as $\Delta x \to 0$, Δu tends to zero too, and, consequently, $\alpha \to 0$. Substituting the expression for Δu into that for Δy, we obtain

$$\begin{aligned}\Delta y &= F'(u)du + F'(u)\beta\,\Delta x + \alpha\,du + \alpha\beta\,\Delta x \\ &= F'(u)du + F'(u)\,\beta\,\Delta x + \alpha f'(x)\,\Delta x + \alpha\beta\,\Delta x \\ &= F'(u)du + (F'(u)\beta + \alpha f'(x) + \alpha\beta)\,\Delta x.\end{aligned}$$

The first term $F'(u)du = F'(u)f'(x)\Delta x$ is proportional to Δx. The brackets of the second term include an infinitesimal since the three terms in brackets are infinitesimals. Consequently, the whole second term

$$(F'(u)\beta + \alpha f'(x) + \alpha\beta)\,\Delta x$$

is an infinitesimal which tends to zero appreciably faster than Δx. Consequently, $F'(u)\,du$ is a differential of the function $y = F(u)$.

This simple formula for the differential of a function of a function means that the form of the notation of a differential does not depend on whether an independent variable or a function of another variable is under the sign of the function. This property of a differential is known as *invariance*.

Example 1. Knowing that $d(x^2) = 2x\,dx$, find $d(x^4)$ representing x^4 as $(x^2)^2$.

Solution. $d(x^2)^2 = 2x^2 d(x^2) = 2x^2 \cdot 2x\,dx = 4x^3\,dx$.

Note how the expression $F'(u)$ must be understood. First we calculate the derivative of the function $F(u)$ as if u is an independent variable. Then we substitute the expression for u as a function of x into the result. Thus

$$F'(u) = F'(f(x)).$$

The formula for the differential immediately yields a formula for the derivative of a function of a function, namely, if $y = F(u) = F(f(x))$, then

$$y' = \frac{dy}{dx} = F'(u)\frac{du}{dx} = F'(u)\,u' = F'(f(x))\,f'(x).$$

The formula for differentiation of a function of a function

$$(F(f(x)))' = F'(f(x))\,f'(x)$$

is very significant for differential and integral calculus.

EXERCISES TO CHAPTER 1

1. Construct a part of the graph of the function $y = x^2$ on the interval [0.5, 1.5]. On [0.5, 1.5] mark points at 0.1 intervals. Accept 5 cm to be a scale unit. Then construct a part of the graph of the function $y = x^2$ on the interval [0.95, 1.05]. Take points at 0.01 intervals and accept 50 cm to be a scale unit (do not try to represent the coordinate axes in the figure). Isolate the part of the first graph which is shown magnified on the second graph.

Tabulate the values of the function $y = 2x - 1$ on the interval [0.95, 1.05] at 0.01 intervals. Follow the variation of the difference $x^2 - (2x - 1)$ at the tabulated points with the approach to $x = 1$.

Make sure that in the second drawing there is no difference between the graphs of the function $y = x^2$ and the straight line $y = 2x - 1$.

2. Construct a part of the graph of the function $y = 1/x$ on the interval [0.5, 2]. Mark points at 0.1 intervals. Accept 5 cm to be a scale unit. Then construct a part of the graph of the function $y = 1/x$ on the interval [0.95, 1.04]. Mark points at 0.01 intervals. Accept 50 cm to be a scale unit. Carry out calculations with an accuracy of 0.0001.

Tabulate the values of the function $y = -x + 2$ on the interval [0.95, 1.04] at 0.01 intervals. Follow the variation of the difference $\frac{1}{x} - (-x + 2)$ at the tabulated points with the approach to $x + 1$.

Make sure that in the second drawing there is no difference between the graphs of the functions $y = 1/x$ and $y = -x + 2$.

3. Use a calculator with an eight-digit mantissa to verify that already at the points $x = 0.9999$ and $x = 1.0001$ the values of the functions $y = 1/x$ and $y = -x + 2$ coincide within the limits of the accuracy provided by the calculator. This means that working with tke approximate value of x which is close to 1, we can find the values of the function $y = 1/x$ from the formula $y = 2 - x$, i.e. replace the operation of division by a simpler operation of subtraction.

Fill in the table below with the values of the function $y = 1/x$, carrying out the calculations with the use of an approximate formula.

x	0.999983	0.999985	0.999987	1.999989
y				
x	1.000015	1.000017	1.000019	1.00002
y				

4. Beginning with what number will the values of the variable x, which assumes a sequence of values $\frac{1}{4}, \frac{1}{7}, \ldots, \frac{1}{3n+1}, \ldots$, become smaller than 10^{-2}, 10^{-4}? Will they become smaller than c, where c is an arbitrary positive number? Is the sequence of values of the variable x vanishing and the variable infinitesimal?

5. Can the terms of the sequence $y_n = \frac{0.0001n+4}{n+2}$ be at a distance smaller than 10^{-1}, 10^{-3}, 10^{-5} from zero? Can we say that the variable y, which assumes this sequence of values, is infinitesimal?

6. The sequence $x_n = 1/n$ is vanishing. Mark by points on the coordinate line as many terms of the sequence x_n as possible, taking 12 blocks on a page of your copybook as a scale unit. Pay attention to the fact that the distance between the adjacent terms decreases rapidly and they become indistinguishable in the figure. Try to mark on this coordinate line the 24th, the 48th and the 100th term of the sequence.

Pay attention to the fact that the terms of the sequence with large numbers practically merge with zero.

Prove that with any choice of the scale, the terms of the sequence with sufficiently large numbers practically merge with the point representing zero in the figure. Show that for the sequence from Exercise 5 the last assertion would not be valid.

7. Prove that the following sequences are vanishing:

1. $y_n = \frac{100}{n}$. 2. $x_n = -\frac{1000}{n}$.

3. $x_n = -\frac{2}{n^2+4}$. 4. $z_n = \frac{1}{n^2+10}$.

5. $z_n = 2^{-n}$. 6. $y_n = \left(\frac{1}{3}\right)^n$.

7. $a_n = \frac{-n-8}{n^2+n+1}$. 8. $b_n = \frac{-2n+1}{n^2-n+6}$.

9. $x_n = \frac{(-1)^n}{n}$. 10. $x_n = \frac{n-10}{n^2+10}$.

8. Prove that any interval of the form $(-\varepsilon, \varepsilon)$, where $\varepsilon > 0$, contains infinitely many terms of a vanishing sequence and that beyond the limits of this interval there are always a finite number of terms of a vanishing sequence.

9. Prove that any vanishing variable is bounded.

10. Is the variable with the values $x_n = \frac{n^2+2}{n^2+1}$ bounded?

11. We have got a new sequence by multiplying every term of the vanishing sequence y_n by 1000. Is the new sequence vanishing?

12. We have got a new sequence by adding 0.001 to every term of a vanishing sequence. Is the new sequence vanishing?

13. Does one of the variables,

x with the values $1, \frac{1}{3}, \frac{1}{5}, \ldots, \frac{1}{2n-1}, \ldots,$

y with the values $\frac{3}{2}, \frac{3}{5}, \frac{3}{10}, \ldots, \frac{3}{n^2+1}, \ldots,$

tend to zero appreciably faster than the other?

14. We divide every term of a vanishing sequence by 1000. Can we say that the new sequence obtained tends to zero appreciably faster than the initial sequence?

15. Find the numbers of the terms of the sequences x_n, which differ from a by less than 0.01, 0.001; by less than ε, where ε is an arbitrary positive number:

1. $x_n = \frac{n}{n+10}$, $a=1$. 2. $x_n = \frac{3n-4}{3n+100}$, $a=1$.

3. $x_n = \frac{20-6n}{n+1}$, $a=-6$. 4. $x_n = \frac{-5n}{n+1}$, $a=-5$.

16. Find the limit of the sequence x_n from Exercise 15 and use appropriate symbols to write it.

17. Prove that the number 3 is not a limit of the sequence $y_n = \frac{50n}{17n+2}$.

Hint. Show that not a single term of the sequence y_n is closer to 3 than by 0.01.

18. Prove that the number 2 is not a limit of the sequence $x_n = \frac{49n}{24n+6}$.

19. It is given that $\lim x_n = a$. How many terms of the sequence x_n are there outside of the interval $(a-\varepsilon, a+\varepsilon)$, where ε is an arbitrary positive number?

20. Prove that any variable which has a limit is bounded.

21. Will the limit of a convergent sequence vary if we add to it a finite number of the first terms or remove a finite number of the first terms?

22. Prove that the limit of a constant quantity (of a variable whose successive values are equal to the same number) is equal to the constant quantity itself.

23. How will the limit of a convergent sequence vary if one and the same number is added to every term of the sequence?

24. How will the limit of a convergent sequence vary if its every term is multiplied by the same number?

25. Prove that if $x_n \to a$, then $\frac{1}{x_n - a} \to \infty$.

26. Prove that a variable with the values $x_n = n - \frac{1}{n}$ tends to infinity.

27. Prove that if $x_n \to \infty$, then $\frac{x_n - 1}{x_n + 1} \to 1$.

28. Given that $x_n \to 1$, $y_n \to -2$, $a_n \to 5$, find the following limits using theorems on limits and the theorem on the limit of a constant quantity:

1. $\lim\limits_{n\to\infty}(x_n+y_n+a_n)$. 2. $\lim\limits_{n\to\infty}(a_n-x_n-y_n)$.

3. $\lim\limits_{n\to\infty}\dfrac{a_n}{y_n}$. 4. $\lim\limits_{n\to\infty}\dfrac{x_n}{a_n}$.

5. $\lim\limits_{n\to\infty}(3x_n-y_n\cdot a_n)$. 6. $\lim\limits_{n\to\infty}(8a_n-x_n\cdot y_n)$.

7. $\lim\limits_{n\to\infty}\dfrac{2x_n+y_n}{a_n}$. 8. $\lim\limits_{n\to\infty}\dfrac{2a_n+5y_n}{x_n}$.

29. Find the following limits:

1. $\lim\limits_{n\to\infty}\dfrac{(n+1)^2+(n+2)^2}{(n+3)^2}$. 2. $\lim\limits_{n\to\infty}\dfrac{3n-4}{n^2+4}$.

3. $\lim\limits_{n\to\infty}(\sqrt{4n^2+5n+6}-2n)$. 4. $\lim\limits_{n\to\infty}(3n-\sqrt{9n^2-6n+2})$.

30. x_n is a sequence of values of the variable x. We derive a formula for the nth term of the corresponding sequence of values of the function $y = x^2$. The formula has the form $y_n = x_n^2$.

Let x_n converge to the number 3.

Find $\lim\limits_{n\to\infty} y_n$. How can the result obtained be formulated in terms of the limit of a function?

31. Find the following limits:

1. $\lim\limits_{x\to 1}(2x+3)$. 2. $\lim\limits_{x\to -1}(-3-8x)$. 3. $\lim\limits_{x\to 3}\dfrac{x^2-x+1}{2x+1}$.

4. $\lim\limits_{x\to 3}\dfrac{3x-4}{x^2-x-2}$. 5. $\lim\limits_{x\to 3}\dfrac{x^2-9}{x-3}$.

6. $\lim\limits_{x\to -1}\dfrac{1-x^2}{x+1}$.

32. Prove that the sum of two functions, which are continuous at a point, is continuous at that point.

33. Prove that the product of two or several functions, which are continuous at a point, is continuous at that point.

34. Prove that any integral rational function is continuous.

35. Prove that any fractional rational function is continuous at every point of the domain of its definition.

36. Prove that if $x \to a$ and $a \geqslant 0$, then $\sqrt{x} \to \sqrt{a}$. Formulate the result obtained in terms of the continuity of a function.

37. Find the following limits using the theorems on continuity proved in Exercises 32-36:

1. $\lim\limits_{x\to 2}\dfrac{3x-4-x^2}{x+3}$. 2. $\lim\limits_{x\to -1}\dfrac{9-x}{x^2+x+1}$.

3. $\lim\limits_{x\to 4} \dfrac{x^2-16}{x-4}$. 4. $\lim\limits_{x\to 5} \dfrac{x^2-4x-5}{x-5}$.

5. $\lim\limits_{x\to 1} \dfrac{\sqrt{x+3}-x-1}{x^2-1}$. 6. $\lim\limits_{x\to 2} \dfrac{\sqrt{x+1}+x-3}{x^2-4}$.

38. Many circuits are intended for conversion of one (input) value to another (output) value according to a certain law. Assume that one

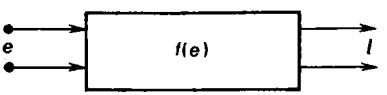


Fig. 1.14

of the devices of this kind (see Fig. 1.14) converts the input signal e into an output quantity I according to the law $I = f(e)$, where $f(e) = 9 + 25e^2$.

(a) What must the input signal be for the output signal differ but little from 10?

(b) What will the value of the output signal be if the input signal differs but little from 2?

(c) What property of the quadratic function $y = ax^2 + b$ allows us to give a substantiated answer to the preceding questions?

39. There are so-called threshold circuits which convert a continuously varying signal into a discontinuously varying output quantity.

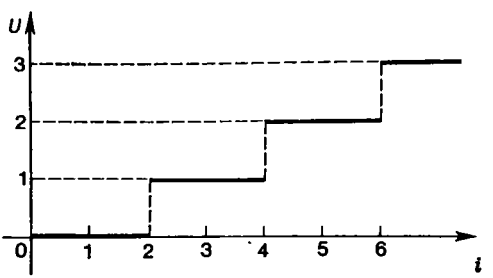


Fig. 1.15

We are given a relationship between the input signal i and the output signal U of one of the circuits of this kind:

$$U(i) = \begin{cases} 0 & \text{for } i < 2, \\ 1 & \text{for } 2 < i < 4, \\ 2 & \text{for } 4 < i < 6, \\ 3 & \text{for } i > 6. \end{cases}$$

Figure 1.15 shows the graph of conversion of the signal i into the signal U.

(a) What can we expect at the output of the circuit if the input signal varies but little about the value 2? 4? 6?

(b) What can we expect at the output if the input signal varies but little about 1? 3? 5?

(c) How can we substantiate the answers to the preceding questions using the properties of the function $y = U(t)$?

40. Find the derivative of each of the following functions:

1. $y = x^3$. 2. $y = 3 - 4x$. 3. $y = \frac{x}{x^2+1}$.

4. $y = \frac{x^2}{x-2}$. 5. $y = \sqrt{x^3}$. 6. $y = \sqrt{5x-3}$.

41. Prove that the function $y = |x|$ does not have a derivative at the point $x = 0$.

42. Prove that the derivative of the linear function $y = kx + b$ is equal to the angular coefficient. Compare the statement proved with the main principle of differential calculus.

43. Set up an equation of a tangent to the graph of each of the following functions at a point with abscissa x_0 or ordinate y_0:

1. $y = x^2$, $x_0 = 1$. 2. $y = x^2$, $x_0 = -1$.

3. $y = \frac{x}{x^2+1}$, $x_0 = 1$. 4. $y = \frac{x^2}{x-2}$, $x_0 = 4$.

5. $y = \frac{1}{x}$, $y_0 = 2$. 6. $y = \frac{1}{x}$, $y_0 = -1$.

7. $y = \sqrt{x}$, $y_0 = \frac{1}{2}$. 8. $y = \sqrt{x}$, $y_0 = 1$.

44. Write the equation of a tangent to the graph of the function $y = \frac{x^2}{x-2}$ at a point with abscissa $x = 3$. Using the equation of the tangent, calculate the approximate value of this function in the vicinity of the point $x = 3$:

1. y (3.001). 2. y (3.0002). 3. y (2.9998). 4. y (2.999).

45. Find the point of tangency and set up an equation of the tangent to the parabola $y = x^2$ if it is known that the tangent passes through the point (2, 3) which does not lie on the parabola.

46. Figure 1.16 shows the graph of the function $y = f(x)$. Find the physical meaning of the average rate of variation of the function on an interval and of the derivative of the function at a point and find the units of their measurement if:

1. The graph shows the path traversed versus time. The path (S) in metres is plotted along the axis of ordinates and the time (t) in minutes along the abscissa axis.

2. The graph shows the process of heating a solid. The temperature (T) in degrees centigrade is plotted along the axis of ordinates and the time (t) in seconds along the abscissa axis.

3. The graph shows the mass of the fuel spent versus the path traversed. The mass (m) of the fuel in grammes is plotted along the axis

of ordinates and the path traversed (S) in kilometres along the axis of abscissas.

4. The graph shows the mass of a part of a nonhomogeneous bar versus the length of that part of the bar. The mass (m) of the part of the bar in grammes is plotted along the axis of ordinates and the length (l) of that part of the bar in centimetres along the abscissa axis (Fig. 1.17).

5. The graph shows the slope of a hill. The altitude (h) above the zero mark, in metres, is plotted along the axis of ordinates and the

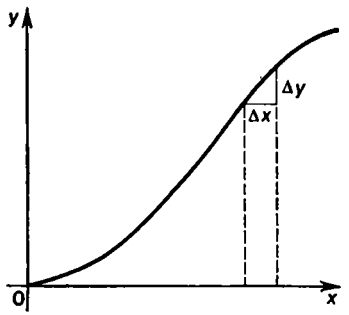


Fig. 1.16

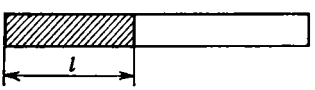


Fig. 1.17

distance (x) along the perpendicular to the zero horizontal, in metres, along the abscissa axis.

6. The graph shows the output voltage of an amplifier versus its input voltage. The output voltage (V) in volts is plotted along the axis of ordinates and the input voltage (e) in volts along the abscissa axis.

47. Use a sheet of millimetre squared paper to construct the graph of a function proceeding from the tabular data, connecting the points by a continuous line:

x	0	0.5	1	2	3	4	5	6	7	8	9
y	0	0.2	0.8	2.9	4.5	5	4.3	2.7	2	1.5	0

The scale is 1 to 10 mm.

Performing necessary constructions, measurements and calculations, find the rate of variation of the function at each of the following points of the domain of definition:

1. $x = 1$. 2. $x = 3$. 3. $x = 6$. 4. $x = 8$.

At which of the points is the rate of increase (decrease) of the function the greatest?

48. A particle moves along a coordinate line according to the law $x(t) = t^2 - 8t - 3$, where x is the coordinate of the particle in metres and t is time in seconds.

Find:

1. The initial position of the particle x (0).
2. The velocity of the particle at an arbitrary moment $v(t)$.
3. The initial velocity of the particle v (0).
4. The acceleration.
5. The moment t_0 of stop (instantaneous), by solving the equation $v(t_0) = 0$.
6. The position of the particle at the moment of the instantaneous stop.
7. The position of the particle ten seconds after the motion begins.
8. The path traversed by the particle during the time from $t = 0$ to $t = 10$.
9. The path traversed by the particle during the time from $t = 6$ to $t = 10$.
10. Construct the graph of motion of the particle where the ordinate is the path $x(t)$ and the abscissa is time t.
11. Write the formula for the path traversed by the partical during the time from $t = 0$ to $t \leqslant t_0$.
12. The way of calculating the path traversed by the particle during the time from $t = 0$ to $t \geqslant t_0$.

49. A particle A moves along the abscissa axis according to the law $x(t) = 4t + 8$ and a point B moves along the axis of ordinates according to the law $x(t) = 3t + 6$. What is the law of variation of the distance between the particles. Find the velocity with which the particles recede from each other.

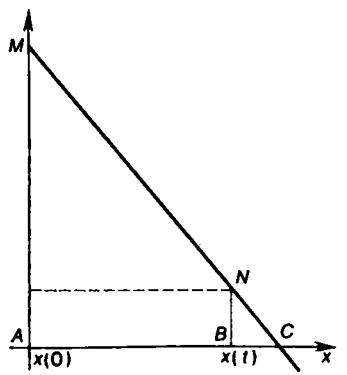


Fig. 1.18

50. A man whose height is 1.8 m walks under a street light with a constant speed of 5 km/h. Set up a formula which will describe the change in the length of the man's shadow from the moment when he was exactly under the light (the length of the shadow was zero) to the moment t. What is the rate of change of the shadow?

The light is at the height of 9.8 m. See Fig. 1.18, where $AB = x(t)$ is the distance from the projection of the light on the road, $BC = l(t)$ is the required length of the shadow, $AM = 9.8$ m and $BN = 1.8$ m.

51. Find the differential of the function $y = x^2 + 1$ considering it to be the principal part of the increment proportional to Δx.

52. Find the differential of the function $y = \sqrt{x}$ from the definition of the differential.

53. Find the differential of the function $y = x^6 = (x^2)^3 = (x^3)^2$, being given that $d(x^2) = 2x\,dx$, $d(x^3) = 3x^2\,dx$.

54. Find the differential of the function $\sqrt{x^2+1}$, being given that $d\sqrt{x} = \dfrac{dx}{2\sqrt{x}}$ and $d(x^2+1) = 2x\,dx$.

Chapter 2

TECHNIQUE OF DIFFERENTIATION

In Chapter 1 we carried out differentiation of some functions proceeding from the definitions of the derivative and the differential. In this chapter we present some rules and formulas which make it possible to carry out differentiation without the use of definitions. These simple rules and formulas give grounds for calling this branch of mathematical analysis *differential calculus.*

2.1. Differentiation of the Results of Arithmetic Operations

Theorem 1. *The derivative of the sum of two functions is equal to the sum of the derivatives of these functions. Correspondingly, the differential of a sum is equal to the sum of differentials.*

Proof. Assume that $u = f_1(x)$ and $v = f_2(x)$ are functions of x and $y = y(x) = u + v = f_1(x) + f_2(x)$. We assign an increment Δx to the value of the variable x and assume that Δu, Δv and Δy are the respective increments of the functions u, v and y. Then

$$\Delta u = f_1(x + \Delta x) - f_1(x), \quad \Delta v = f_2(x + \Delta x) - f_2(x),$$

$$\Delta y = f_1(x + \Delta x) + f_2(x + \Delta x) - (f_1(x) + f_2(x))$$

$$= f_1(x + \Delta x) - f_1(x) + f_2(x + \Delta x) - f_2(x) = \Delta u + \Delta v.$$

Consequently, $\frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} + \frac{\Delta v}{\Delta x}$ and

$$\lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{\Delta u}{\Delta x} + \lim_{\Delta x \to 0} \frac{\Delta v}{\Delta x},$$

so that $y' = u' + v'$, and this is what we wished to prove. Multiplying both sides of this relation by dx, we obtain $y'\,dx = u'\,dx + v'\,dx$, i.e. $dy = du + dv$.

It is clear that we have actually proved Theorem 1 in the following stronger formulation.

If the functions $u = f_1(x)$ and $v = f_2(x)$ have derivatives at a point, then their sum has a derivative at the same point and this derivative is equal to the sum of the derivatives of the summands.

Theorem 2. *The derivative of the product $y = cu$ of a constant c by a function $u = f(x)$ is equal to the product of the same constant by the derivative of the function u.*

It can be formalized as $(cu)' = c \cdot u'$.

For differentials $d(cu) = c\,du$.

Proof. We set $y = cu = cf(x)$. Then $\Delta y = cf(x+\Delta x) - cf(x) = c[f(x+\Delta x) - f(x)] = c\,\Delta u$. Consequently, $\frac{\Delta y}{\Delta x} = c\frac{\Delta u}{\Delta x}$ and $\lim\limits_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = c \lim\limits_{\Delta x \to 0} \frac{\Delta u}{\Delta x}$. And this means that $y' = cu'$. Multiplying by dx, we get $dy = c\,du$.

Theorem 3. *The derivative of the product of two functions is equal to the product of the derivative of the first factor by the second summed up with the product of the first factor by the derivative of the second.*

This rather complicated formulation is a verbal expression of the following simple formula:

$$(uv)' = u'v + uv'.$$

Similarly, for differentials we have

$$d(uv) = (du)v + u\,(dv).$$

Proof. Let $u = f_1(x)$, $v = f_2(x)$ and $y = uv = f_1(x)\,f_2(x)$. Assume that Δx is the increment of the independent variable x, and Δu, Δv and Δy are the increments of the functions u, v and y respectively. Then

$$\Delta y = (u + \Delta u)(v + \Delta v) - uv = uv + (\Delta u)v + u(\Delta v) + (\Delta u)(\Delta v) - uv = (\Delta u)v + u(\Delta v) + \Delta u \Delta v.$$

Division by Δx yields

$$\frac{\Delta y}{\Delta x} = \frac{\Delta u}{\Delta x} v + u \frac{\Delta v}{\Delta x} + \frac{\Delta u}{\Delta x} \cdot \Delta v.$$

The limiting process on $\Delta x \to 0$ yields

$$y' = (uv)' = u'v + uv' + u' \lim_{\Delta x \to 0} \Delta v = u'v + uv' + u' \cdot 0 = u'v + uv',$$

and this is what we wished to prove.

Multiplying by dx, we get a corresponding formula for the differential:

$$d(uv) = (du)v + u(dv).$$

When we prove theorems 2 and 3, we again assume the existence of derivatives of the functions u and v and on this assumption we prove the existence (and not only derive a formula) of the derivatives of the functions uv and cu.

Theorem 4. *The derivative of a constant (i.e. a function assuming the same value for all values of the independent variable) is equal to zero.*

Proof. Let $y = f(x) = c$, where c is a constant. Then

$$\Delta y = f(x + \Delta x) - f(x) = c - c = 0, \quad \frac{\Delta y}{\Delta x} = 0,$$

and

$$y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 0.$$

Theorem 5. *The derivative of an independent variable is equal to* 1.

Proof. Let $y = x$. Then $\Delta y = \Delta x$, $\frac{\Delta y}{\Delta x} = 1$ and $y' = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = 1$.

Theorem 6. *Let n be a natural number. The derivative of the power function $y = x^n$ is equal to nx^{n-1}.*

Proof. Assume that $n = 2$ (we considered this case in 1.8 proceeding from the definition of the derivative), i.e. $y = x^2 = x \cdot x$. Then, according to Theorem 3, y' is equal to

$$x'x + xx' = 1 \cdot x + x \cdot 1 = 2x.$$

Assume now that $n = 3$, i.e. $y = x^3 = x^2 x$. Then

$$y' = (x^2)'x + x^2 x' = 2xx + x^2 \cdot 1 = 3x^2.$$

By analogy, we can present arguments for $n = 4, 5, \ldots$.

To prove the formula for any n, it is natural to use induction (since the method of induction usually replaces a chain

of arguments of the same type). Assume that we have derived a formula for the exponent $n-1$: $(x^{n-1})' = (n-1)\,x^{n-2}$. Then, for $y = x^n$ we have

$$y = x^n = x^{n-1}x,$$

$$y' = (x^{n-1})'\,x + x^{n-1}x' = (n-1)\,x^{n-2}x + x^{n-1}\cdot 1$$

$$= (n-1)x^{n-1} + x^{n-1} = nx^{n-1}.$$

We have proved the theorem (it holds true for $n = 2$, and, consequently, for $n = 3$ as well; it holds true for $n = 3$, and, consequently, for $n = 4$ as well, and so on).

It is clear that the formula

$$d(x^n) = nx^{n-1}\,dx$$

holds true for the differential of a power with a natural exponent.

Theorem 7. *The derivative of a fraction at a point at which the fraction has sense (i.e. the denominator does not vanish) is equal to a fraction whose numerator is equal to the difference of the products of the derivative of the numerator by the denominator and the numerator by the derivative of the denominator and the denominator is equal to the square of the denominator of the initial fraction.*

This formulation is the verbal expression of the formula

$$\left(\frac{u}{v}\right)' = \frac{u'v - uv'}{v^2}.$$

Proof. Assume that $y = u/v$, where $u = f_1(x)$, $v = f_2(x)$ and $v \neq 0$ at the point being considered. Assume furthermore that Δx is the increment of the independent variable x and Δu, Δv and Δy are the respective increments of the functions u, v and y. Then

$$\Delta y = \frac{u+\Delta u}{v+\Delta v} - \frac{u}{v} = \frac{(u+\Delta u)\,v - u\,(v+\Delta v)}{v\,(v+\Delta v)}$$

$$= \frac{uv + (\Delta u)\,v - uv - u\,(\Delta v)}{v\,(v+\Delta v)} = \frac{(\Delta u)\,v - u\,(\Delta v)}{v\,(v+\Delta v)}.$$

Division by Δx and a limiting process as $\Delta x \to 0$ yields

$$y' = \left(\frac{u}{v}\right)' = \lim_{\Delta x \to 0} \frac{\frac{\Delta u}{\Delta x} \cdot v - u \frac{\Delta v}{\Delta x}}{v\left(v + \frac{\Delta v}{\Delta x} \cdot \Delta x\right)}$$

$$= \frac{\lim\limits_{\Delta x \to 0} \frac{\Delta u}{\Delta x} \cdot v - u \lim\limits_{\Delta x \to 0} \frac{\Delta v}{\Delta x}}{v\left(v + \lim\limits_{\Delta x \to 0} \frac{\Delta v}{\Delta x} \cdot \Delta x\right)} = \frac{u'v - uv'}{v(v + v' \cdot 0)}$$

$$= \frac{u'v - uv'}{v^2}.$$

Remark. In these arguments there is no reason for apprehension in connection with the transfer of $v + \Delta v$ into the denominator. We assume, of course, that the function v has a derivative at the point being considered and is consequently continuous. Therefore, $\Delta v \to 0$ as $\Delta x \to 0$. The value of v is nonzero and, consequently, $v + \Delta v$ becomes nonzero for a sufficiently small Δx.

Theorem 8. *The formula* $(x^n)' = nx^{n-1}$ *remains valid for a negative integral exponent.*

Proof. Let $y = x^n$, where n is a negative integer. We set $n = -k$, where k is now a natural number. Then $y = x^{-k} = \frac{1}{x^k}$. According to Theorem 7 we have

$$y' = \frac{1' \cdot x^k - 1 \cdot (x^k)'}{(x^k)^2} = \frac{-kx^{k-1}}{x^{2k}} = -k \frac{1}{x^{k+1}} = -kx^{-k-1} = nx^{n-1}.$$

Theorems 1-6 give means for calculating the derivative of any integral rational function. Indeed, an integral rational function can be reduced to the form of a polynomial, i.e. of the sum of powers of an independent variable with some coefficients. It also follows from these theorems (in the stronger formulation) that an integral rational function has a derivative at any point. The derivative again proves to be an integral rational function and, consequently, is continuous everywhere. Therefore, an integral rational function is a smooth function.

Theorem 7 yields a rule of differentiation of a fractional rational function. The derivative exists at all points except for the points at which the denominator vanishes. The derivative again proves to be a fractional rational function whose

denominator vanishes at the same points as the denominator of the initial function. Therefore a fractional rational function turns out to be piecewise smooth and the open intervals of smoothness are divided by points at which the denominator vanishes.

Let us now consider some examples of application of the theorems we have proved.

Example 1. Find the derivative and the differential of the function

$$y = x^5 + 5x^4 - 10x^2 + 3.$$

Solution. We apply theorems 1, 2, 4 and 6:

$$y' = 5x^4 + 20x^3 - 20x, \quad dy = (5x^4 + 20x^3 - 20x)\,dx.$$

Example 2. Find the derivative and the differential of the function

$$y = \frac{x^3}{x^2+1}.$$

Solution.

$$y' = \frac{3x^2(x^2+1) - x^3 \cdot 2x}{(x^2+1)^2} = \frac{x^4+3x^2}{(x^2+1)^2},$$

$$dy = \frac{x^4+3x^2}{(x^2+1)^2}\,dx.$$

Example 3. Find the derivative of the function

$$y = \frac{x^2+1}{x^3}.$$

Solution 1.

$$y' = \frac{x^3 \cdot 2x - (x^2+1)\,3x^2}{x^6} = \frac{-x^4-3x^2}{x^6} = \frac{-x^2(x^2+3)}{x^6}$$

$$= -\frac{x^2+3}{x^4}.$$

Solution 2 (with the use of Theorem 8).

$$y = (x^2+1)x^{-3},$$

$$y' = 2x \cdot x^{-3} + (x^2+1)(-3x^{-4})$$

$$= \frac{2}{x^2} - \frac{3x^2+3}{x^4} = \frac{-x^2-3}{x^4} = -\frac{x^2+3}{x^4}.$$

Solution 3.

$$y = \frac{x^2+1}{x^3} = \frac{x^2}{x^3} + \frac{1}{x^3} = x^{-1} + x^{-3},$$

$$y' = -x^{-2} - 3x^{-4} = -\frac{1}{x^2} - \frac{3}{x^4} = -\frac{x^2+3}{x^4}.$$

Example 4. Find the derivative of the function

$$y = (x^2 + x + 1)^2.$$

Solution 1.

$$y = x^4 + 2x^3 + 3x^2 + 2x + 1,$$

$$y' = 4x^3 + 6x^2 + 6x + 2.$$

Solution 2 (with the use of calculation of a function of a function).

$$y = (x^2 + x + 1)^2,$$

$$y' = 2(x^2 + x + 1)\,(x^2+x+1)' = 2(x^2 + x+1)\,(2x+1)$$
$$= 4x^3 + 6x^2 + 6x + 2.$$

Example 5. Find the derivative of the function

$$y = \frac{1}{(x^2+1)^4}.$$

Solution.

$$y' = -4(x^2+1)^{-5} \cdot (x^2+1)' = -\frac{4}{(x^2+1)^5} \cdot 2x = -\frac{8x}{(x^2+1)^5}.$$

2.2. Differentiation of a Logarithmic Function

Let $y = \log_a x$. Then

$$y' = \lim_{\Delta x \to 0} \frac{\log_a (x+\Delta x) - \log_a x}{\Delta x} = \lim_{\Delta x \to 0} \frac{\log_a \dfrac{x+\Delta x}{x}}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\log_a \left(1 + \dfrac{\Delta x}{x}\right)}{\Delta x}, \quad x > 0.$$

We set $\frac{\Delta x}{x} = h$. Then $h \to 0$ as $\Delta x \to 0$ and $\Delta x = xh$. Therefore

$$y' = \lim_{h \to 0} \frac{1}{x} \frac{\log_a (1+h)}{h} = \frac{1}{x} \lim_{h \to 0} \log_a (1+h)^{1/h}$$

$$= \frac{1}{x} \log_a \left(\lim_{h \to 0} (1+h)^{1/h} \right).$$

We set $x=1$ and obtain

$$y'_{x=1} = \log_a\left(\lim_{h\to 0}(1+h)^{1/h}\right).$$

Thus, for any point x, $x>0$, we have

$$y' = \frac{1}{x}\cdot y'_{x=1}.$$

The value $y'_{x=1}$ is the slope of the tangent at the point $x = 1$, i.e. at the point at which the graph of the logarithmic function cuts the abscissa axis. Since all logarithmic functions are proportional (i.e. differ by a constant factor), there are functions among them whose graphs cut the abscissa axis at various angles (except for the right angle). In particular, there is a logarithmic function the slope of whose tangent at the point of intersection of the graph with the abscissa axis is equal to 1. A logarithmic function of this kind is known as a *natural function* and the logarithms corresponding to it are called *natural logarithms*. The base of the system of natural logarithms is the number $e = \lim\limits_{h\to 0}(1+h)^{1/h}$, known as a *Napierian number*. This is an irrational, and even transcendental, number, meaning that it is not a root of any algebraic equation with integral coefficients. It has been calculated with more than 1000 decimal digits and is equal to 2.718281828459045 with an accuracy of 10^{-15}.

The natural logarithm of the number x is designated as $\ln x$.

Thus a natural logarithmic function has $\frac{1}{x}$ as its derivative and its differential is $\frac{dx}{x}$. The derivative of a logarithmic function $\log_a x$ is $\frac{1}{x}\log_a e$.

Let us consider some examples.

Example 1. Find the derivative of the function $y = \ln(x^2+1)$.

Solution. We must use here the rule of calculation of the derivative of a function:

$$y' = \frac{1}{x^2+1}(x^2+1)' = \frac{2x}{x^2+1}.$$

Formally, it is simpler to begin with the differential:

$$dy = \frac{d(x^2+1)}{x^2+1} = \frac{2x\,dx}{x^2+1} \quad \text{and} \quad y' = \frac{dy}{dx} = \frac{2x}{x^2+1}.$$

Example 2. Find the derivative of the function $y = (\ln x)^2$.

Solution.

$$dy = 2\ln x \cdot d\ln x = 2\ln x \frac{dx}{x},$$

$$y' = \frac{dy}{dx} = \frac{2\ln x}{x}.$$

Example 3. Find the derivative of the function $y = \frac{\ln x}{x}$.

Solution.

$$y' = \frac{x(\ln x)' - \ln x \cdot x'}{x^2} = \frac{x \cdot \frac{1}{x} - \ln x}{x^2} = \frac{1 - \ln x}{x^2}.$$

◀ 2.3. Proof of the Existence of a Limit of the Function $(1 + h)^{1/h}$ as $h \to 0$

In the preceding section we assumed without proof, on the basis of a geometrically obvious fact that the graph of a logarithmic function cuts the abscissa axis at a definite angle, that the limit $(1 + h)^{1/h}$ must exist as $h \to 0$.

We shall carry out here a more rigorous proof of the existence of a limit.

Lemma. *Assume that* $\alpha > -1$, $\alpha \neq 0$, n *is a natural number. Then* $(1 + \alpha)^n > 1 + n\alpha$.

Proof. Let $n = 2$. We have $(1 + \alpha)^2 = 1 + 2\alpha + \alpha^2 > 1 + 2\alpha$ since $\alpha^2 > 0$. Furthermore, $(1 + \alpha)^3 = (1 + \alpha)^2 \times (1 + \alpha) > (1 + 2\alpha)(1 + \alpha)$ since $1 + \alpha > 0$ and $(1 + \alpha)^2 > 1 + 2\alpha$. But $(1 + 2\alpha)(1 + \alpha) = 1 + 3\alpha + 2\alpha^2 > 1 + 3\alpha$. Thus $(1 + \alpha)^3 > 1 + 3\alpha$. We can reason in an analogous way for $n = 4, 5, \ldots$. Therefore we must use induction. We assume as already proved that $(1 + \alpha)^{n-1} > 1 + (n - 1)\alpha$. Then $(1 + \alpha)^n = (1 + \alpha)^{n-1}(1 + \alpha) > (1 + (n - 1)\alpha)(1 + \alpha)$ since $(1 + \alpha)^{n-1} > 1 + (n - 1)\alpha$ and $1 + \alpha > 0$. But $(1 + (n - 1)\alpha)(1 + \alpha) = 1 + (n - 1)\alpha + \alpha + (n - 1)\alpha^2 = 1 + n\alpha + (n - 1)\alpha^2 > 1 + n\alpha$ since $\alpha^2 > 0$. And this completes the proof of the lemma.

Theorem 1. *Assume that* $n \to \infty$ *running over a natural series of numbers. Then there is a limit* $\lim\limits_{n\to\infty}\left(1+\frac{1}{n}\right)^n$.

Proof. We set $\left(1+\frac{1}{n}\right)^n = u_n$. We shall show first of all that u_n increases with n. With this aim in view, we consider

$$\frac{u_{n+1}}{u_n} = \frac{\left(1+\frac{1}{n+1}\right)^{n+1}}{\left(1+\frac{1}{n}\right)^n} = \left(1+\frac{1}{n}\right)\frac{\left(1+\frac{1}{n+1}\right)^{n+1}}{\left(1+\frac{1}{n}\right)^{n+1}}$$

$$= \left(1+\frac{1}{n}\right)\left(\frac{1+\frac{1}{n+1}}{1+\frac{1}{n}}\right)^{n+1} = \left(1+\frac{1}{n}\right)\left(\frac{n(n+2)}{(n+1)^2}\right)^{n+1}$$

$$= \left(1+\frac{1}{n}\right)\left(\frac{n^2+2n+1-1}{(n+1)^2}\right)^{n+1}$$

$$= \left(1+\frac{1}{n}\right)\left(1-\frac{1}{(n+1)^2}\right)^{n+1}.$$

Now we use the lemma and obtain

$$\left(1+\frac{1}{n}\right)\left(1-\frac{1}{(n+1)^2}\right)^{n+1} > \left(1+\frac{1}{n}\right)\left(1-\frac{n+1}{(n+1)^2}\right)$$

$$= \frac{n+1}{n}\cdot\frac{n}{n+1} = 1.$$

Thus $\frac{u_{n+1}}{u_n} > 1$, i.e. $u_{n+1} > u_n$ and this holds true for all n.

Then we introduce into consideration a new sequence of numbers: $v_n = \left(1+\frac{1}{n}\right)^{n+1}$. We shall show that v_n decreases with an increase in n. With this aim in view we consider

$$\frac{v_n}{v_{n+1}} = \frac{\left(1+\frac{1}{n}\right)^{n+1}}{\left(1+\frac{1}{n+1}\right)^{n+2}} = \frac{1}{\left(1+\frac{1}{n}\right)}\cdot\frac{\left(1+\frac{1}{n}\right)^{n+2}}{\left(1+\frac{1}{n+1}\right)^{n+2}}$$

$$= \frac{n}{n+1}\left(\frac{1+\frac{1}{n}}{1+\frac{1}{n+1}}\right)^{n+2} = \frac{n}{n+1}\left(\frac{(n+1)^2}{n(n+2)}\right)^{n+2}$$

$$= \frac{n}{n+1}\left(\frac{n^2+2n+1}{n^2+2n}\right)^{n+2}$$
$$= \frac{n}{n+1}\left(1+\frac{1}{n^2+2n}\right)^{n+2}.$$

Now we use the lemma and obtain

$$\frac{v_n}{v_{n+1}} > \frac{n}{n+1}\left(1+\frac{n+2}{n^2+2n}\right) = \frac{n}{n+1}\left(1+\frac{1}{n}\right) = 1.$$

Thus $\frac{v_n}{v_{n+1}} > 1$ and, consequently, $v_{n+1} < v_n$.

We have found that the sequence $u_1, u_2, \ldots, u_n, \ldots$ is increasing and the sequence $v_1, v_2, \ldots, v_n, \ldots$ is decreasing. Furthermore,

$$v_n = \left(1+\frac{1}{n}\right)^{n+1} = \left(1+\frac{1}{n}\right)\left(1+\frac{1}{n}\right)^n = \left(1+\frac{1}{n}\right)u_n,$$

and, consequently, $v_n > u_n$.

For $n = 1, 2, \ldots$ the intervals $[u_n, v_n]$ are nested intervals since u_n is increasing and v_n is decreasing. They constitute a nest since $v_n - u_n = \frac{1}{n}u_n < \frac{1}{n}v_n < \frac{1}{n}v_1 = \frac{4}{n}$ and $\frac{4}{n} \to 0$ as $n \to \infty$. Therefore, $v_n - u_n \to 0$.

A unique real number (we designate it as e) which belongs to all intervals of the nest $[u_n, v_n]$ is the limit of the sequence $u_n = \left(1+\frac{1}{n}\right)^n$. Indeed, $e - u_n > 0$ and $|u_n - e| = e - u_n < v_n - u_n < \frac{4}{n} \to 0$. Therefore $|u_n - e| \to 0$ and $e = \lim\limits_{n\to\infty} u_n$. The sequence $v_n = \left(1+\frac{1}{n}\right)^{n+1}$ converges to the same limit.

Theorem 2. *Let h run over a sequence of positive values $h_1, h_2, \ldots, h_k, \ldots$, which converges to zero. Then*

$$(1+h_k)^{1/h_k} \to e.$$

Proof. We can certainly assume that all h_k are nonzero since otherwise the expression $(1+h_k)^{1/h_k}$ would be meaningless. A "large" positive number $1/h_k$ is included between two natural numbers:

$$n_k \leqslant 1/h_k < n_k + 1.$$

It is clear that $n_k \to \infty$ as $h_k \to 0$. Furthermore, the inequality $\frac{1}{n_k+1} < h_k \leqslant \frac{1}{n_k}$ is fulfilled for h_k. From the inequalities for $1/h_k$ and h_k we infer that

$$(1+h_k)^{1/h_k} < \left(1+\frac{1}{n_k}\right)^{n_k+1} = v_{n_k}$$

and

$$\begin{aligned}(1+h_k)^{1/h_k} &> \left(1+\frac{1}{n_k+1}\right)^{n_k} \\ &= \left(1+\frac{1}{n_k+1}\right)^{n_k+1} \cdot \left(1+\frac{1}{n_k+1}\right)^{-1} \\ &= u_{n_k+1}\left(1+\frac{1}{n_k+1}\right)^{-1}.\end{aligned}$$

Thus $(1+h_k)^{1/h_k}$ is included between the numbers v_{n_k} and $u_{n_k+1} \cdot \left(1+\frac{1}{n_k+1}\right)^{-1}$ which evidently have the same limit e. Consequently, $\left(1+\frac{1}{h_k}\right)^{h_k}$ tends to the same limit.

Theorem 3. *Let h run over a sequence of negative values h_1, h_2, . . ., h_k converging to zero. Then $(1+h_k)^{1/h_k} \to e$.*

Proof. We set $h_k = -g_k$, where g_k is a sequence of positive numbers converging to zero. Then

$$\begin{aligned}(1+h_k)^{1/h_k} &= (1-g_k)^{-1/g_k} = \left(\frac{1}{1-g_k}\right)^{1/g_k} \\ &= \left(1+\frac{g_k}{1-g_k}\right)^{1/g_k}.\end{aligned}$$

We set $\frac{g_k}{1-g_k} = t_k$. It is clear that $t_k \to 0$ as $g_k \to 0$, and the values t_k become positive as soon as the values g_k become smaller than 1.

From the relation $\frac{g_k}{1-g_k} = t_k$ we find that

$$g_k = \frac{t_k}{1+t_k}, \quad \frac{1}{g_k} = \frac{1+t_k}{t_k} = 1 + \frac{1}{t_k}.$$

Thus $(1+h_k)^{1/h_k} = (1+t_k)^{1/t_k+1} = (1+t_k)^{1/t_k}(1+t_k)$. As $t_k \to 0$, the first factor converges to e and the second to 1.

Thus $\lim (1 + h)^{1/h}$ is again equal to e as h approaches zero from the side of the negative values. We have proved the theorem.

It follows from theorems 2 and 3 that $\lim_{h \to 0} (1 + h)^{1/h}$ exists and is equal to e since if h assumes positive as well as negative values, which tend to zero, the function $(1 + h)^{1/h}$ infinitely approaches e both for positive and for negative values of h. ▶

2.4. Differentiation of an Exponential Function

Let $y = a^x$ for $a > 0$ and $a \neq 1$. Then $\ln y = x \ln a$. Differentiating both sides of this relation, we get $\frac{dy}{y} = \ln a \cdot dx$, whence we have $dy = y \ln a \, dx = a^x \ln a \, dx$. Dividing by dx, we obtain

$$(a^x)' = \frac{dy}{dx} = a^x \ln a.$$

If $a = e$, i.e. the exponential function in question is a natural exponential function, then the formula for the derivative assumes an especially simple form:

$$(e^x)' = e^x.$$

The formula for the differential is $de^x = e^x \, dx$.

The simplicity of these formulas is one of the reasons for expressing the exponential function a^x with any base in terms of a natural exponential function, i.e. in terms of an exponential function with a base e. This is done as follows. We set $\ln a = k$. Then, in accordance with the definition of a logarithm, $a = e^k$, whence $a^x = (e^k)^x = e^{kx}$. The formula for the derivative of an exponential function derived above assumes the form $(e^{kx})' = ke^{kx}$. For a positive k the function e^{kx} evidently increases with x. The differentiation formula shows that the rate of increase of the function e^{kx} is proportional to its value at the point being considered. For a negative k the function e^{kx} decreases and the rate of its decrease is proportional to its value. It is clear that the same properties are inherent in the rates of increase (decrease) of a function of the form ce^{kx} for $c > 0$.

Let us consider some examples.

Example 1. Find the derivative of the function $y = 10^x$.

Answer. $y' = 10^x \ln 10$.

Example 2. Find the derivative of the function $y = e^{x^2}$.

Solution.

$$dy = de^{x^2} = e^{x^2} d(x^2) = e^{x^2} 2x\,dx \quad \text{and} \quad y' = \frac{dy}{dx} = e^{x^2} \cdot 2x.$$

Example 3. Find the derivative of the function $y = (x^2 + 1)e^x$.

Solution.

$$y' = (x^2 + 1)' \cdot e^x + (x^2 + 1)(e^x)'$$
$$= 2x \cdot e^x + (x^2 + 1)e^x = (x^2 + 2x + 1)e^x = (x + 1)^2 e^x.$$

Example 4. Find the derivative of the function $y = x^x$, $x > 0$.

Solution.

$$y = (e^{\ln x})^x = e^{x \ln x},$$

$$dy = e^{x \ln x} d(x \ln x) = x^x ((dx) \cdot \ln x + x d \ln x)$$
$$= x^x \left(\ln x \cdot dx + x \frac{dx}{x} \right) = x^x (\ln x + 1)\,dx,$$

$$y' = \frac{dy}{dx} = x^x (\ln x + 1).$$

2.5. Differentiation of a Power Function

Let $y = x^a$. For integral (positive and negative) values of a we have derived a differentiation formula somewhat earlier: $(x^a)' = ax^{a-1}$. Now we shall set up a differentiation formula for a power function $y = x^a$ for any real exponent a. When a is not an integer, we must assume that $x > 0$.

Taking a natural logarithm of both sides of the relation $y = x^a$, we get $\ln y = a \ln x$. Equating the differentials, we obtain $\frac{dy}{y} = a\frac{dx}{x}$, whence it follows that $dy = a\frac{y}{x}dx = ax^{a-1}\,dx$. Dividing by dx, we get

$$y' = \frac{dy}{dx} = ax^{a-1}.$$

Thus the formula for the derivative of a power function with any real exponent has the same form as for an integral exponent.

It is easy to see that when $a > 1$ a derivative exists for $x = 0$ as well and is equal to 0.

Let us consider some examples showing how to calculate derivatives and differentials.

Example 1. $y = x^{\sqrt{2}}$.

Answer. $y' = \sqrt{2}\, x^{\sqrt{2}-1}$.

Example 2. $y = \sqrt{x^2+1}$.

Solution.

$$y = (x^2+1)^{1/2},$$

$$dy = \frac{1}{2}(x^2+1)^{-1/2}\, d(x^2+1) = \frac{1}{2\sqrt{x^2+1}} 2x\, dx = \frac{x}{\sqrt{x^2+1}}\, dx,$$

$$y' = \frac{dy}{dx} = \frac{x}{\sqrt{x^2+1}}.$$

Example 3. $y = \dfrac{x}{\sqrt{x^2+1}}$.

Solution.

$$y = x(x^2+1)^{-1/2},$$

$$dy = (dx)(x^2+1)^{-1/2} + x \cdot \left(-\frac{1}{2}\right)(x^2+1)^{-3/2}\, d(x^2+1)$$

$$= \frac{dx}{\sqrt{x^2+1}} - \frac{x}{2\sqrt{(x^2+1)^3}} \cdot 2x\, dx$$

$$= \frac{dx}{\sqrt{(x^2+1)^3}}(x^2+1-x^2) = \frac{dx}{\sqrt{(x^2+1)^3}},$$

$$y' = \frac{dy}{dx} = \frac{1}{\sqrt{(x^2+1)^3}}.$$

2.6. Differentiation of Functions Defined by Equations

If the graph of the relation $F(x, y) = 0$ between two variables x and y has the form of a continuous line and the point (x_0, y_0) lies on this graph and is not a point of self-intersection of the graph, then, in a sufficiently small neighbourhood of this point, we can regard one of the variables as a function of the other. And it turns out that we can calculate the derivative of this function. To do this, we must take a differential of the function $F(x, y)$, equate it to zero and find $\frac{dy}{dx}$ $\left(\text{or } \frac{dx}{dy}\right)$ from the relation obtained.

Example 1. From the equation of a circle $x^2+y^2=r^2$ we find that $x\,dx+y\,dy=0$, whence $\frac{dy}{dx}=-\frac{x}{y}$. Here the equation defines two smooth functions $y=\sqrt{r^2-x^2}$ and $y=-\sqrt{r^2-x^2}$. For both functions we have $y'=-x/y$. We can also verify this by means of a direct calculation of the derivative, proceeding from an explicit formula.

Example 2. The point (1, 1) lies on a curve with an equation $x^3+y^3=2xy$. Find the slope of the tangent to the curve at this point.

Solution. The differentiation of the relation yields $3x^2\,dx+3y^2\,dy=2y\,dx+2x\,dy$, whence

$$\frac{dy}{dx}=\frac{3x^2-2y}{2x-3y^2}.$$

Substituting the coordinates of the point into the right-hand side of the last relation, we get the answer:

$$y'_{\substack{x=1\\y=1}}=\frac{3-2}{2-3}=-1.$$

EXERCISES TO CHAPTER 2

1. Find the derivative and the differential of each of the following functions:

1. $y=2x^7-3x^5+4x^3-x-10$. 2. $y=3x^6-x^4-3x^2+7$.
3. $u=4t^4+2t^2$. 4. $v=\frac{1}{5}t^5-\frac{1}{3}t^3$. 5. $f=\frac{1}{z^3}-\frac{1}{z}$.
6. $g=-\frac{1}{\tau^4}+\frac{1}{\tau^2}-\frac{1}{\tau}$. 7. $h=t^4-\frac{1}{t^4}$. 8. $f=\frac{1}{t^3}-t^3$.
9. $y=\frac{x}{x-3}$. 10. $y=\frac{x-7}{x+5}$. 11. $y=\frac{t^2-1}{t^2+1}$.
12. $y=\frac{t^3+2}{t^3-t}$. 13. $f=\frac{x^2-3x-1}{x^2+x+4}$. 14. $h=\frac{5x-x^2-3}{4-x-x^2}$.
15. $g=(3u+7)^5$. 16. $f=(7-5u)^4$.
17. $S=2\pi r^2+4\pi lr+k$, where π, l, k are constants and r is the argument.
18. $V=\frac{4}{3}\pi r^3-\frac{2}{3}\pi hr^2+k$, where π, h, k are constants and r is the argument.
19. $F=2\pi(4-l)^2-\pi(H-3l)^2$, where π, H are constants and l is the argument.

20. $H=(3-ax-x^2)(2+bx)$, where a, b are constants and x is the argument.

21. $P=\dfrac{a}{(\alpha-4)^2}-\dfrac{b}{(\alpha+3)^3}$, where a, b are constants and α is the argument.

22. $G=\dfrac{\alpha}{(t^2-1)^3}+\dfrac{\beta}{(t^3-8)^2}$, where α, β are constants and t is the argument.

2. $y=(2x-1)(3x-2)(4x-3)$. Find: $y'(x)$, $y'(1)$, $y'\left(\frac{1}{2}\right)$, $y'\left(\frac{2}{3}\right)$, $y'\left(\frac{3}{4}\right)$.

3. $y=x^3(5x-1)(1-2x)$. Find: $y'(x)$, $y'(0)$, $y'\left(\frac{1}{2}\right)$.

4. Prove that $(u\cdot v\cdot w)'=u'\cdot v\cdot w+u\cdot v'\cdot w+u\cdot v\cdot w'$.

5. Find x for which the tangents to the graphs of the functions

$$y=x^3-x^2-x+4, \quad y=\frac{2}{3}x^3+2x-3$$

are parallel.

6. Find the points at which the tangent to the graph of the function

$$f=\frac{4}{3}x^3+4x^2-20x-3$$

is parallel to the abscissa axis.

7. Write the equation of the tangent to the graph of the function $y=f(x)$ at the point with abscissa x_0:

1. $y=\dfrac{x^3+2x^2-x-5}{x^4}$, $\quad x_0=1$.

2. $y=\dfrac{x^2-3x-1}{x^2-3}$, $\quad x_0=-1$.

8. The law of a rectilinear motion of a particle is described by a piecewise-defined function $s(t)$:

$$s(t)=\begin{cases} 4t & \text{for } 0\leqslant t\leqslant 3, \\ -t^2+10t-9 & \text{for } 3\leqslant t\leqslant 5, \\ 16 & \text{for } t\geqslant 5, \end{cases}$$

where s is the path in metres and t is time in seconds. Find the velocity and the acceleration at the moments $t=2$, $t=4$, $t=6$. Show the law of motion by graphical means.

9. The law of variation of the velocity of a moving particle is described by a piecewise-defined function $v(t)$:

$$v(t)=\begin{cases} 5 & \text{for } 0\leqslant t\leqslant 4 \\ 2t-3 & \text{for } 4\leqslant t\leqslant 6, \\ 21-2t & \text{for } 6\leqslant t\leqslant 10.5, \end{cases}$$

where v is velocity in m/s and t is time in seconds. Find the acceleration of motion at moments $t = 2$, $t = 5$, $t = 8$. Show the law of variation of the velocity and the acceleration by graphical means.

10. Find the derivative of each of the following functions:

1. $\ln(3x-7)$. 2. $\ln(5+2x)$. 3. $\ln\left(\frac{1}{2}x^2+x\right)$.

4. $\ln\left(x-\frac{1}{3}x^3\right)$. 5. $\log_2 x$. 6. $\log_{\frac{1}{3}} x$. 7. $\log^3 x$.

8. $\frac{1}{\log^2 x}$. 9. $x^2 \ln x$. 10. $(x-4)^3 \ln x$.

11. $\ln(x+1)(x+2)(x+3)$. 12. $\ln\frac{x-2}{(x+3)(x-4)}$.

13. $\log_{\frac{1}{2}} \frac{x^2-x}{x+7}$. 14. $\log_3 (x+x^2)(x-2)$. 15. $\frac{\ln x}{x^2}$.

16. $\frac{\ln x}{x-2}$. 17. $\ln(-x)$. 18. $\ln(5-2x)$.

11. Compare the rate of variation of the following functions at the same points:

$$f_1(x) = \ln x,\ f_2(x) = \ln 2x,\ f_3(x) = \ln \frac{x}{10}.$$

12. Write the equation of the tangent to the graph of each of the following functions at the point with abscissa x_0:

1. $f(x) = \ln x$, $x_0 = 1$.
2. $f(x) = \ln x^2$, $x_0 = e$ (e is Napier's number).
3. $f(x) = \log x$, $x_0 = 2$.

13. Compare the rates of variation of the following functions at the same points:

$$\varphi_1(x) = \log_2 x,\quad \varphi_2(x) = \log_4 x,\quad \varphi_3(x) = \log_8 x.$$

14. $f(x) = \ln(x-1)(x-3)(x-5)$. Find $f'(2)$, $f'(6)$, $f'(4)$. First find the domain of definition of the function.

15. Find the sign of the derivative of each of the following functions:

1. $y = \ln(3-2x)$. 2. $y = \log_{0.5}(2x-5)$.

16. Find the derivative of each of the following functions:

1. 3^x, 2. $\left(\frac{1}{2}\right)^x$. 3. e^{x^2-x}. 4. e^{1-2x}. 5. $t^2 e^t$.

6. te^{t^2}. 7. $\frac{z}{e^{z^2}}$. 8. $\frac{z^2}{e^z}$. 9. $\left(\left(\frac{1}{2}\right)^{1-x}\right)^2$.

10. $(3^{x-1})^{-2}$. 11. $4 \cdot 2^{3t-2}$. 12. $\frac{1}{9} \cdot 3^{2t+2}$.

13. $(5x+1)^x$. 14. x^{3x-2}.

17. Write the equation of the tangent to the graph of the function $f(x)$ at the point with abscissa x_0:

1. $f(x) = e^x$, $x_0 = 0$. 2. $f(x) = (1/e)^x$, $x_0 = 0$.

3. $f(x) = 2^{3x-4}$, $x_0 = 1$. 4. $f(x) = 3^{2x+3}$, $x_0 = -1$.

18. Find the point beginning with which the rate of variation of the function $y = e^x$ becomes higher than the rate of variation of the function $y = 2x + 1$.

19. Find the derivative of each of the following functions:

1. $x^{-1.3}$. 2. $x^{2/3}$. 3. $\sqrt{7x-1}$. 4. $\sqrt{5-3x}$.

5. $\dfrac{x-2}{\sqrt{4-3x}}$. 6. $\dfrac{2-3x}{\sqrt{2x+1}}$. 7. $\sqrt[3]{-5t+2}$.

8. $\sqrt[3]{1+4x}$, 9. $\sqrt[3]{(u^2-u)^2}$. 10. $\sqrt[3]{(3x+x^2)^4}$.

20. Write the equation of the tangent to the graph of the function $f(x)$ at the point with abscissa x_0:

1. $f(x) = \sqrt{x}$, $x_0 = 4$. 2. $f(x) = \sqrt{4-x}$, $x_0 = 0$.

3. $f(x) = x^{3/2}$, $x_0 = 4$. 4. $f(x) = (-x)^{5/3}$, $x_0 = -1/8$.

21. Is the rate of variation of the function $y = \sqrt{x}$ higher or lower than that of the function $y = x^2$ at the point $x_0 = \dfrac{1}{16}$? at the point $x_0 = \dfrac{1}{2}$? or at the point $x_0 = 1$?

22. At what moment do the rates of variation of the functions $y = x^2$ and $y = \sqrt{x}$ become equal?

23. Find the derivative of the function $y(x)$ without expressing it in the explicit form in terms of x:

1. $3y - 4x = 7$. 2. $2x + 4y - 1 = 0$. 3. $x \cdot y = 2$.

4. $y(x - 5) = -4$. 5. $xy - 2y + x - 7 = 0$.

6. $2xy - 6x + y - 11 = 0$.

24. Find the slope of the tangent to the curve defined by the equation $F(x, y) = 0$ at the point (x_0, y_0):

1. $y^3 - 2xy + x^2 - 1 = 0$, $(1, 1)$.

2. $-x^3 + xy^2 - y^2 - 5 = 0$, $(-2, 1)$.

Chapter 3

SOME APPLICATIONS OF DIFFERENTIAL CALCULUS

3.1. Tests for Increase and Decrease of Functions

The function $f(x)$ defined on an interval is *increasing* if the larger values of the argument are associated with the larger values of the function, i.e. if $f(x_2) > f(x_1)$ as soon as $x_2 > x_1$. Correspondingly, the function is *decreasing* if $x_2 > x_1$ implies that $f(x_2) < f(x_1)$. Increasing and decreasing functions are known as *monotonic functions*. A function

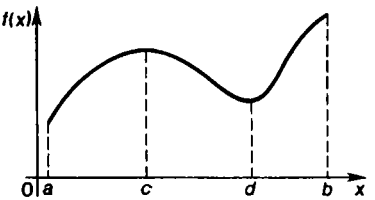


Fig. 3.1

is *piecewise-monotonic* if any finite interval included in its domain consists of several intervals on each of which the function is monotonic (Fig. 3.1).

The main principle of differential calculus yields simple tests for increase and decrease of differentiable functions (i.e. functions which have a derivative).

Let the function $f(x)$ be increasing on an interval. Then its graph is a curve rising with the movement from left to right (Fig. 3.2). Therefore a small section of a tangent which almost coincides with the portion of the graph adjacent to the point of tangency also rises, or, at least, is horizontal. Consequently, the slope of the tangent at any point of the curve (i.e. the value of the derivative) is larger than or equal to zero. We have thus established (at the level of strictness of the "main principle") the validity of the following theorem.

Theorem 1. *If a differentiable function increases on an interval, then its derivative has nonnegative values at all points of the interval.*

The following theorem, which is almost converse to Theorem 1, is also true.

Theorem 2. *If the derivative $f'(x)$ of the function $f(x)$, defined on an interval, assumes positive values at all points of the interval, then the function $f(x)$ increases on that interval.*

Proof. Indeed, the inequality $f'(x) > 0$ at all points of the interval shows that a tangent at any point of the graph is a

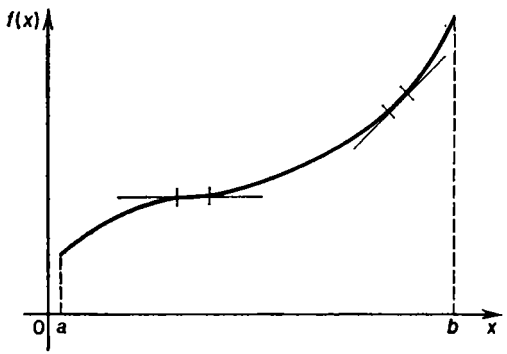


Fig. 3.2

rising straight line and, consequently, any, sufficiently small, portion of the graph is a rising line. It can be understood by intuition that the graph as a whole is a rising line and, consequently, the function $f(x)$ increases.

◀ We have said that Theorem 2 is almost converse to Theorem 1. This "almost" is due to the fact that in the conclusion of Theorem 1 the vanishing of the derivative is admitted whereas the hypothesis of Theorem 2 rules out this possibility.

We can easily strengthen Theorem 2 by admitting the vanishing of the derivative at a finite number of points on the interval. Indeed, excluding these points, we divide the whole interval into a finite number of open intervals in each of which the derivative assumes only positive values, and, consequently, the function increases in each of these intervals. By virtue of continuity, the graph is composed of a finite number of rising lines and, consequently, the whole line is rising. So the function increases throughout the interval. ▶

We can combine the two theorems in the following formulation.

For a differentiable function to increase on an interval, it is necessary that its derivative should assume nonnegative values and sufficient that all the values of the derivative should be positive.

The tests for a decrease of functions are quite similar to those of an increase, namely, the following theorems hold true.

Theorem 3. *If a differentiable function decreases on an interval, then its derivative has nonpositive values at all points of the interval.*

Theorem 4. *If the derivative $f'(x)$ of the function $f(x)$, defined on an interval, assumes negative values at all points of the interval, then the function $f(x)$ decreases on the interval.*

To prove these theorems, we can reason as we did when we proved theorems 1 and 2. But we can do a simpler thing, namely, we can apply theorems 1 and 2 to the function $-f(x)$ since it is clear that if $f(x)$ decreases, then $-f(x)$ increases and $-f'(x) = (-f(x))'$.

Example 1. Find the intervals of monotonicity of the function $y = \frac{x}{x^2+1}$ first by elementary means and then with the use of the sufficient conditions for increase and decrease of a function in an interval.

Solution 1. We assume that $x_2 > x_1$ and consider the difference

$$y(x_2) - y(x_1) = \frac{x_2}{x_2^2+1} - \frac{x_1}{x_1^2+1} = \frac{x_2 \cdot x_1^2 + x_2 - x_1 \cdot x_2^2 - x_1}{(x_2^2+1)(x_1^2+1)}$$

$$= \frac{(x_2 - x_1) - x_2 \cdot x_1 (x_2 - x_1)}{(x_2^2+1)(x_1^2+1)},$$

$$= \frac{(x_2 - x_1)(1 - x_1 \cdot x_2)}{(x_2^2+1)(x_1^2+1)}.$$

The first factor $(x_2 - x_1)$ in the numerator is positive since $x_2 > x_1$ and the denominator of the fraction is obviously also larger than zero. Thus the sign of the fraction coincides with the sign of the expression $1 - x_1 \cdot x_2$. On the interval $[-1, 1]$ the product $x_1 \cdot x_2$ is either negative or larger than zero but smaller than 1. Thus, on the interval $[-1, 1]$ the expression $1 - x_1 \cdot x_2$ and, hence, the difference $y(x_2) - y(x_1)$ assume positive values, and, consequently, $y(x)$ increases on the

interval $[-1, 1]$ since a larger value of the function corresponds to a larger value of the argument.

On the intervals $[1, \infty)$ and $(-\infty, -1]$ the product $x_1 \cdot x_2 > 1$ and, consequently, the expression $1 - x_1 \cdot x_2$ and, hence, the difference $y(x_2) - y(x_1)$ are negative, and, consequently, in the intervals $[1, \infty)$ and $(-\infty, -1]$ the function increases.

Solution 2. We shall solve the problem using the derivative We have

$$y'(x) = \frac{x^2+1-2x^2}{(x^2+1)^2} = \frac{1-x^2}{(x^2+1)^2}.$$

We test the derivative for sign. Since the denominator of the fraction is positive for all x, the sign of the fraction

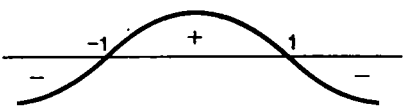


Fig. 3.3

coincides with that of the quadratic function $1 - x^2$. Figure 3.3 shows a conventional distribution of the signs of the derivative along the number axis. Applying the sufficient conditions for monotonicity of a function on an interval, we infer that $y(x)$ increases on $[-1, 1]$ and decreases on $(\infty, -1]$ and on $[1, \infty)$. Note that the endpoints of the intervals (the points $x = -1$ and $x = 1$) are included both into the interval on which the function decreases and into the interval on which the function increases.

This example vividly shows the advantage of the method of solution based on differential calculus. In many cases technical difficulties do not allow us to solve a problem of this kind by elementary means at all.

Example 2. Find the intervals of monotonicity of the function $y = \frac{1}{5}x^5 - \frac{2}{3}x^3 + x + 3$.

Solution. We find the derivative of this function:

$$y' = x^4 - 2x^2 + 1 = (x^2 - 1)^2 = (x - 1)^2 (x + 1)^2,$$

and test it for sign. Figure 3.4 shows the signs of the derivative. Throughout the interval $(-\infty, \infty)$, except for the

points $x = -1$ and $x = 1$, the derivative is greater than zero and at the points $x = -1$ and $x = 1$ the derivative is zero. By virtue of continuity, the function $y = \frac{1}{5}x^5 - \frac{2}{3}x^3 + x + 3$ increases throughout the interval $(-\infty, \infty)$.

At the points $x = -1$ and $x = 1$ the increase of the function is terminated "for an instant" but as soon as the

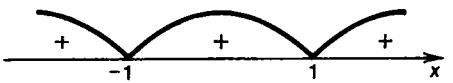


Fig. 3.4

function passes these points, it immediately begins to increase.

Example 3. Construct the graph of the function $y = \frac{1}{5}x^5 - \frac{2}{3}x^3 + x + 3$ considered in the preceding example. In the same drawing depict the graph of the derivative of this function. Write the equations of the tangents to the graph of the function at the points $x = -1$ and $x = 1$.

Note that in the vicinity of the points of tangency the curve lies on different sides of the tangent. Compare this fact with the character of the variation of the derivative of the function in the vicinity of the points of tangency and find one more point in whose vicinity the graph of the given function lies on different sides of the tangent.

◀ 3.2. Refinement of the Proofs of the Theorems on the Increase and Decrease of Functions

In 3.1 we carried out the proofs of theorems 1-4 on the tests for the increase and decrease of functions in a graphic form without adhering to the strictness of arguments typical of mathematics. In this section we present more rigorous arguments.

Lemma 1 (Rolle's Theorem). *If the function $f(x)$ is differentiable at all points of the interval $[a, b]$ and vanishes at the endpoints of this interval, i.e. $f(a) = f(b) = 0$, then there is a point within the interval such that $f'(c) = 0$.*

(In other words, *between two roots of a differentiable function there is at least one root of the derivative.*)

Proof. If the function $f(x)$ does not assume positive or negative values, i.e. is zero throughout the interval, then its derivative is zero at all points of the interval, so that the lemma is true on this interval. Let us assume now that the function $f(x)$ assumes positive values. Then there is a point c on the interval at which the function assumes the greatest value (we regard this as an obvious fact). The point c does not coincide with the endpoints a and b of the interval. We shall show that $f'(c) = 0$. To do this, we consider the ratio $\frac{f(c+\Delta x)-f(c)}{\Delta x}$. It is clear that $f(c+\Delta x)-f(c) \leqslant 0$ for all values of Δx, both positive and negative, since $f(c)$ is the greatest value of the function $f(x)$. Therefore $\frac{f(c+\Delta x)-f(c)}{\Delta x} \leqslant 0$ for $\Delta x > 0$ and $\frac{f(c+\Delta x)-f(c)}{\Delta x} \geqslant 0$ for $\Delta x < 0$. By virtue of differentiability of the function $f(x)$, there is a limit of the ratio $\frac{f(c+\Delta x)-f(c)}{\Delta x}$ for Δx tending to zero over any sequence of values, in particular, the limit is the same for a sequence of positive values of $\Delta x \to 0$ and for a sequence of negative values of $\Delta x \to 0$. Consequently,

$$f'(c) = \lim_{\Delta x \to 0} \frac{f(c+\Delta x)-f(c)}{\Delta x} \leqslant 0$$

and

$$f'(c) = \lim_{\Delta x \to 0} \frac{f(c+\Delta x)-f(c)}{\Delta x} \geqslant 0,$$

whence $f'(c) = 0$, and this is what we wished to prove.

It remains to consider the case when $f(x)$ does not assume positive values and is different from a function which is zero everywhere. In this case $f(x)$ assumes negative values and the function $-f(x)$ assumes positive values. For the function $-f(x)$ the conditions of the lemma are fulfilled. Consequently, there is a point c such that $(-f(x))'_{x=c} = 0$, i.e. $-f'(c) = 0$ and, therefore, $f'(c) = 0$. This completes the proof of the lemma.

Lemma 2 (the first mean value theorem for derivatives). *Let $f(x)$ be differentiable on the interval $[a, b]$. Then, within this interval, there is a point c such that* $f'(c) = \frac{f(b)-f(a)}{b-a}$.

Proof. We choose a linear function $l(x) = kx + m$ such that its values at the endpoints of the interval coincide with the values of the function $f(x)$. We can do this since we can draw a straight line through any two points (Fig. 3.5). Then the conditions of Rolle's theorem are fulfilled for the difference $g(x) = f(x) - l(x)$. Therefore there is a point c such that $g'(c) = 0$. But $g'(c) = f'(c) - l'(c) = f'(c) - k$. Thus we have $f'(c) = k$ for this point. Let us now find k. For this purpose, we write the requirements, which the function $l(x)$ must satisfy, in the form of an equation, namely,

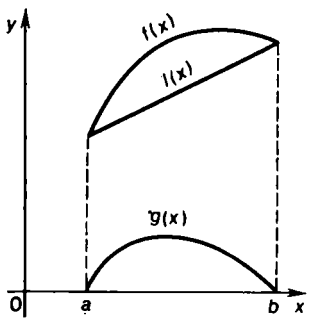


Fig. 3.5

$$f(a) = ka + m, \quad f(b) = kb + m.$$

From this system we find that $k = \frac{f(b) - f(a)}{b - a}$. Consequently, $f(c) = \frac{f(b) - f(a)}{b - a}$, and this is what we wanted to prove.

Note that $f(b) - f(a)$ is the increment of the function $f(x)$, which corresponds to the increment $b - a$ of the argument x when we pass from $x = a$ to $x = b$. Thus the theorem states that the ratio of the increment of a differentiable function to the increment of the argument on an interval is equal to the value of the derivative at a point within that interval. This theorem is known as the *mean value theorem* or the *law of the mean.*

We can now present more rigorous proofs of theorems 1-4.

Theorem 1. *If a differentiable function increases on an interval, then its derivative has nonnegative values at all points of the interval.*

Proof. If the function increases, then

$$f(x_0 + \Delta x) > f(x_0) \text{ for } \Delta x > 0$$

and

$$f(x_0 + \Delta x) < f(x_0) \text{ for } \Delta x < 0.$$

In both cases $\frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} > 0$. Consequently,

$$f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x} \geqslant 0.$$

This is true for all points x_0 of the interval. We have proved the theorem.

Theorem 2. *If the derivative $f'(x)$ of the function $f(x)$, defined on an interval, assumes positive values at all points of the interval, then the function $f(x)$ increases.*

Proof. According to Lemma 2, for any x_1 and x_2 of the interval (let $x_1 < x_2$), there holds a relation

$$\frac{f(x_2)-f(x_1)}{x_2-x_1} = f'(c)$$

for a certain c of the interval (x_1, x_2). It follows from the hypothesis that $f'(c) > 0$ and so $\frac{f(x_2)-f(x_1)}{x_2-x_1} > 0$, whence $f(x_2) - f(x_1) > 0$ since $x_2 - x_1 > 0$. Thus $f(x_2) > f(x_1)$, and this means that the function $f(x)$ increases.

As we saw in 3.1, theorems 3 and 4, yielding the tests for the decrease of functions, can be easily derived from theorems 1 and 2.

There are other tests for the increase (decrease) of a smooth (differentiable) function besides that based on the mean value theorem. We shall discuss one of the techniques which we consider to be very significant and instructive from many points of view.

We say that the function $f(x)$ *increases at a point* x_0 if the inequality $f(x) < f(x_0)$ holds true in a sufficiently small neighbourhood of this point when $x < x_0$ and if the inequality $f(x) > f(x_0)$ holds true when $x > x_0$. If x_0 lies at the endpoint of the interval of definition of the function, then one of these inequalities must be satisfied, namely,

$$f(x) > f(x_0) \text{ for } x > x_0$$

for the left endpoint, and

$$f(x) < f(x_0) \text{ for } x < x_0$$

for the right endpoint. If follows from the definition that if $[c, d]$ is a sufficiently small interval containing the point x_0, then $f(c) < f(d)$.

It is clear that if $f(x)$ *increases on an interval, then it increases at all of its points.*

In the same way we define a decrease of a function at a point. It is clear that *if a function decreases on an interval, then it decreases at all of its points.*

Lemma 3. *If a function increases at every point of an interval, then it increases on that interval.*

At first glance the statement of the lemma seems to be obvious. However, if we think it over, it becomes not so obvious because of the difference between the definitions of an increase of a function on an interval and its increase at a point. In the first definition we compare the values of the function at any pair of points of the interval and in the second, we compare the values at a point with the values at the points of its sufficiently small neighbourhood.

Proof of the lemma. Let the function $f(x)$ increase at all points of an interval. Assume that the function does not increase on that interval, i.e. that there are numbers a_1 and b_1 such that $a_1 < b_1$ but $f(a_1) \geqslant f(b_1)$. Let us reduce this assumption to contradiction. Consider a number $c_1 = \frac{a_1+b_1}{2}$. It is clear that $a_1 < c_1 < b_1$. It may happen so that $f(c_1) \geqslant f(b_1)$. In this case we change the designations setting $c_1 = a_2$ and $b_1 = b_2$. But it is not improbable that $f(c_1) < f(b_1)$. Then, so much the more, $f(c_1) < f(a_1)$. In this case we set $a_1 = a_2$, $c_1 = b_2$. In both cases we have constructed an interval $[a_2, b_2]$ nested in the interval $[a_1, b_1]$ which is half as long and such that $f(a_2) \geqslant f(b_2)$. Now we present the same arguments for the numbers a_2, b_2 and $c_2 = \frac{a_2+b_2}{2}$. We get a new interval $[a_3, b_3]$ nested in the interval $[a_2, b_2]$ which is half as long and such that $f(a_3) \geqslant f(b_3)$. Continuing this process, we get a collection of nested intervals.

By virtue of the axiom on continuity, there is a point x_0 which belongs to all these intervals. Thus, for this point there are intervals $[a_n, b_n]$, of an arbitrarily small length, such that $f(a_n) \geqslant f(b_n)$ and $a_n \leqslant x_0 \leqslant b_n$. This contradicts the definition of an increase of a function at the point x_0 (to be more precise, to the obvious corollary of this definition). This contradiction is at variance with the assumption that the function $f(x)$ does not increase throughout the interval.

We have thus proved the lemma.

Let us return to the proofs of theorems 1-4 on an increase and decrease of functions. We shall give a stronger statement of Theorem 1.

Theorem 1′. *If a differentiable function $f(x)$ increases at a point x_0 of an interval, then its derivative has a nonnegative value at this point.*

Proof. Let $f(x)$ increase at the point x_0. This means that we have

$$f(x_0 + \Delta x) - f(x_0) > 0 \text{ for } \Delta x > 0$$

and

$$f(x_0 + \Delta x) - f(x_0) < 0 \text{ for } \Delta x < 0$$

for a sufficiently small $|\Delta x|$. In both cases

$$\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} > 0, \quad f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} \geqslant 0.$$

Theorem 1′ immediately yields Theorem 1 since if a function increases on an interval, then it increases at all of its points.

Theorem 2′. *If the value of the derivative is positive at a point, then the function increases at that point.*

Proof. Assume that $f'(x_0) = \lim_{\Delta x \to 0} \frac{f(x_0+\Delta x)-f(x_0)}{\Delta x} > 0$. Then, for Δx sufficiently small in absolute value, the ratio $\frac{f(x_0+\Delta x)-f(x_0)}{\Delta x}$ exceeds 0. For $\Delta x > 0$ this means that $f(x_0 + \Delta x) - f(x_0) > 0$, i.e. $f(x_0 + \Delta x) > f(x_0)$. For a negative Δx we have $f(x_0 + \Delta x) - f(x_0) < 0$, i. e. $f(x_0 + \Delta x) < f(x_0)$. These inequalities mean that $f(x)$ increases at the point x_0.

As in the case when we used the mean value theorem, we find that Theorem 2′ and Lemma 3 immediately yield the validity of Theorem 2. ▶

3.3. Maxima and Minima of Functions

The point x_0 is a *point of local minimum* for the function $f(x)$ if its value $f(x_0)$ at this point is smaller than all its values in some neighbourhood of this point. The value $f(x_0)$ itself is a *local minimum* of the function $f(x)$. The word "local" emphasizes the fact that the value $f(x_0)$ is the least of the values only in one, maybe small, neighbourhood of the point x_0, and far from the point x_0 the values may be smaller

than $f(x_0)$. For a function defined on an interval the *absolute minimum* is the value of the function which is the least of the values throughout the interval.

In the same way we can define a *point of local maximum* as a point x_0 for which $f(x_0)$ is the greatest of the values in a neighbourhood of the point x_0, a *local maximum of the function* as the value $f(x_0)$ at a point of local maximum and the *absolute maximum* as the greatest value of the function defined on the interval. In Fig. 3.6 the point x_0 is a point of local minimum of the function $f(x)$ and x_1 is a point of local maximum. The absolute minimum and absolute maximum are attained at the endpoints a and b on the interval of definition of the function.

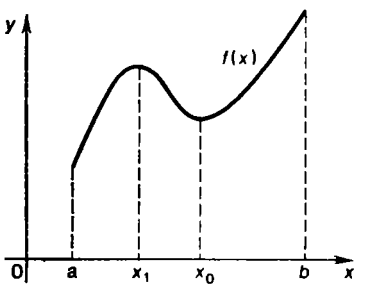


Fig. 3.6

The common name of the maximum and minimum of a function is an *extremum* and the points at which they are attained are *points of extremum*.

Theorem. *If the point x_0 of extremum of the function $f(x)$ lies in the interval of its definition and the derivative $f'(x_0)$ exists at this point, then $f'(x_0) = 0$.*

Proof. If $f'(x_0) > 0$, then the tangent rises from left to right and a portion of the graph of the function, closely adjoining the tangent, rises together with it. Therefore the points of the graph on the left of x_0 lie lower than $f(x_0)$ and the points of the graph on the right of x_0 lie above $f(x_0)$. Consequently, x_0 cannot be either a point of local maximum or a point of local minimum.

Now if $f'(x_0) < 0$, then the tangent, and, together with it, the portion of the graph, go down from left to right. Therefore $f(x) > f(x_0)$ for x which is sufficiently close to x_0 and $x < x_0$, and $f(x) < f(x_0)$ for x which is sufficiently close to x_0 and $x > x_0$. Consequently, here again x_0 cannot be either a point of local minimum or a point of local maximum. The only possibility that remains is $f'(x_0) = 0$, and this proves the theorem.

◀ A somewhat more rigorous proof of the theorem is actually in the proof of Lemma 1 (Rolle's theorem) in 3.2. ▶

For a piecewise monotonic function, every point of local maximum separates the interval of increase of the function from the interval of its decrease (we regard the intervals from left to right). Correspondingly, the points of local minimum separate the intervals of decrease of the function from the intervals of its increase. Therefore the derivative $f'(x)$ of a piecewise monotonic function changes sign from plus to minus when x increases and passes through the point of

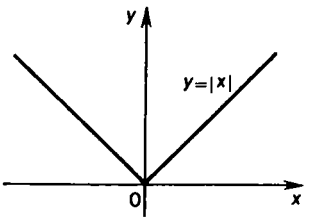


Fig. 3.7

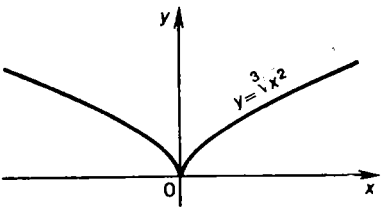


Fig. 3.8

maximum. Correspondingly, $f'(x)$ changes sign from minus to plus when x increases and passes through the point of minimum.

At the point of extremum the derivative of even a "rather good" piecewise monotonic function may not exist. The function $y = |x|$ is a typical example of this. It evidently has a minimum for $x = 0$ but does not have a derivative at this point (Fig. 3.7). The function $y = \sqrt[3]{x^2}$ may serve as another typical example. The point $x = 0$ is a point of minimum for it. For $x \neq 0$ its derivative is $\frac{2}{3\sqrt[3]{x}}$, for $x = 0$ the derivative does not exist, y' tends to infinity as $x \to 0$ (Fig. 3.8). We can easily see that a semicubic parabola with the equation $y^2 = x^3$, turned through 90°, is the graph of the function $y = \sqrt[3]{x^2}$.

When we test functions for an extremum, we must not forget to investigate the points at which the derivative does not exist, provided that there are points of this kind.

For a piecewise monotonic function defined on a closed interval, the endpoints of the interval are always points of local extremum. Namely, if $f(x)$, defined on $[a, b]$ increases when x recedes to the right of the left endpoint a of the

interval, then a is a point of minimum, and if it decreases, then a is a point of maximum. Correspondingly, if $f(x)$ increases when x approaches the right endpoint b of the interval from the left, then b is a point of maximum and if $f(x)$ decreases, then b is a point of minimum.

It is clear that the absolute maximum of a piecewise monotonic function, defined on a closed interval, is among the local maxima and the absolute minimum is among the local minima. For a function defined in an open interval, there may be no local maximum or local minimum.

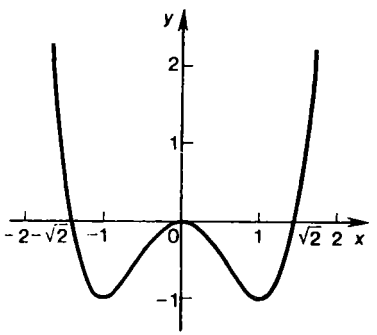


Fig. 3.9

Thus a function equal to $y=1/x$ does not have an absolute maximum or an absolute minimum on the interval $(0, +\infty)$.

Let us consider some examples.

Example 1. For the function $y = x^4 - 2x^2$ find the points of maximum and minimum and the intervals of increase and decrease.

Solution. $y' = 4x^3 - 4x = 4x(x^2 - 1) = 4x(x+1)(x-1)$. At three points, $x = -1$, $x = 0$ and $x = 1$, the derivative vanishes. These points divide the abscissa axis into four parts: $(-\infty, -1)$, $(-1, 0)$, $(0, 1)$ and $(1, \infty)$. On the interval $(-\infty, -1)$ the function y' is negative, on the interval $(-1, 0)$ it is positive, on the interval $(0, 1)$ it is negative and on the interval $(1, +\infty)$ it is positive. Consequently, on $(-\infty, -1)$ y decreases, on $(-1, 0)$ it increases, on $(0, 1)$ it decreases and on $(1, +\infty)$ it increases. Therefore the points $x = -1$ and $x = +1$ are points of minimum with the values of the function -1 and the point $x = 0$ is a point of maximum with the value 0. The graph of the function is presented in Fig. 3.9. As $x \to \pm\infty$, the function $y = x^4\left(1 - \frac{1}{x^2}\right)$ rapidly (approximately as x^4) tends to $+\infty$. The graph cuts the abscissa axis at the points $x = \pm\sqrt{2}$.

Example 2. Do the same as in Example 1 for the function $y = \frac{x}{x^2+1}$.

Solution. $y' = \dfrac{(x^2+1)-x2x}{(x^2+1)^2} = \dfrac{1-x^2}{(x^2+1)^2}$. At two points, $x = \pm 1$, the derivative vanishes. On the intervals $(-\infty, -1)$ and $(1, +\infty)$ the derivative is negative, and on the interval $(-1, 1)$ it is positive. Therefore the point -1 is a point of minimum and the point 1 is a point of maximum.

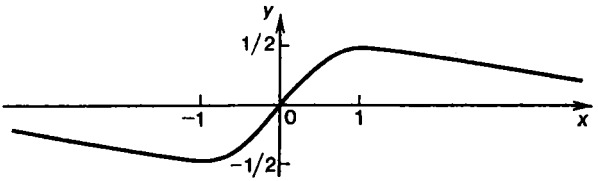


Fig. 3.10

As $x \to \infty$, the function tends to zero from above as $x \to +\infty$ and tends to zero from below as $x \to -\infty$. The graph is shown in Fig. 3.10. The intervals of increase and decrease were defined in Example 1 in 3.1.

Example 3. $y = \dfrac{1-x}{x^2}$.

Solution. $y = x^2 - x^{-1}$, $y' = -2x^{-3} + x^{-2} = \dfrac{x-2}{x^3}$. The point of discontinuity $x = 0$ divides the natural domain of existence into two intervals, $(-\infty, 0)$ and $(0, +\infty)$. On the first interval the derivative is positive and so the function increases. As $x \to -\infty$, the function $y \to 0$ from above, i.e. assuming positive values. As $x \to 0$ from the left, $y \to +\infty$. On the interval $(0, 2)$ the derivative is negative and so the function decreases. On the interval $(2, +\infty)$ the derivative is positive and so the function increases. Therefore, at the point $x = 2$ it has a minimum for the value $-1/4$. As $x \to +\infty$, $y \to 0$ from below, i.e. assuming negative values. The graph once cuts the abscissa axis at the point $x = 1$. Thus the graph has the shape shown in Fig. 3.11.

◀**Example 4.** $y = x^3 + 3\,|\,x\,|$.

Solution. Here we must express the function by two different formulas for the intervals $(-\infty, 0)$ and $[0, +\infty)$. On the first interval $y = x^3 - 3x$, on the second $y = x^3 + 3x$. For $-\infty < x < 0$ we have $y' = 3x^2 - 3$. The derivative vanishes for $x = 1$ and $x = -1$. However, the first point

does not belong to the interval $(-\infty, 0)$. On the interval $(-\infty, -1)$ the derivative is positive, on the interval $(-1, 0)$ it is negative, and therefore -1 is a point of maximum. On the interval $(0, \infty)$ the derivative $y' = 3x^2 + 3 > 0$, and so, on this interval, the function increases.

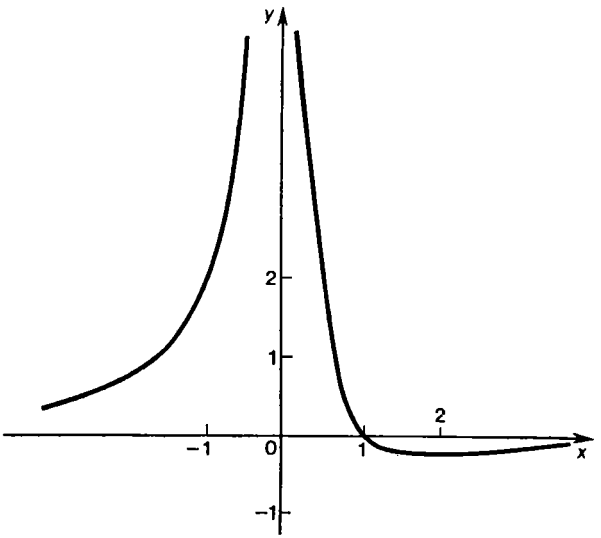


Fig. 3.11

This means that the point $x = 0$, at which the derivative of the function $y = x^3 + 3 \mid x \mid$ does not exist, proves to be a point of minimum. The graph has the shape shown in Fig. 3.12.

Example 5. $y = 2x - 3\sqrt[3]{x^2}$.

Solution. Here $y' = 2 - 3 \cdot \dfrac{2}{3\sqrt[3]{x}} = \dfrac{2(\sqrt[3]{x} - 1)}{\sqrt[3]{x}}$.

The points $x = 0$ (where there is no derivative) and $x = 1$ (where the derivative is 0) are "doubtful" as far as the extremum is concerned. These points divide the whole natural domain $(-\infty, +\infty)$ of definition of the function into three intervals, $(-\infty, 0)$, $(0, 1)$ and $(1, +\infty)$. On the first interval, i.e. for $x < 0$, the derivative is positive and tends to

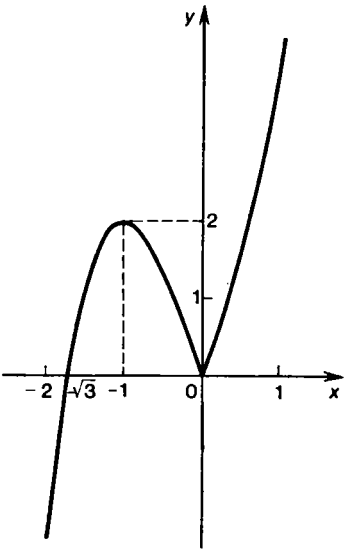


Fig. 3.12

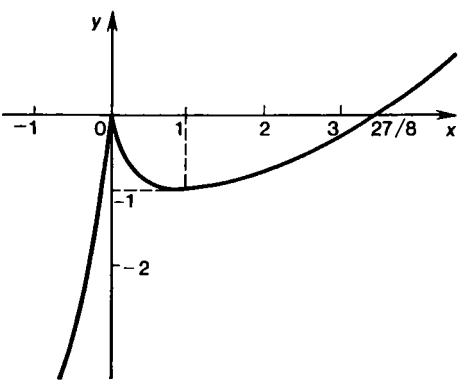


Fig. 3.13

infinity as $x \to 0$ from the left. On the second interval, i.e. for $0 < x < 1$, the derivative is negative and tends to $-\infty$ as $x \to 0$ from the right. Finally, for $x > 1$ the derivative is positive. Consequently, the function increases on $(-\infty, 0)$, decreases on $(0, 1)$ and increases on $(1, \infty)$. The point $x = 0$ is a point of maximum with a "spike" (since the derivative tends to infinity as $x \to 0$) and the point $x = -1$ is an "ordinary" point of minimum. The graph cuts the abscissa axis, at another point beside $x = 0$ and in the interval of the increase of the function $1 < x < \infty$, namely, at $x = 27/8$. Furthermore, $y = 2x\left(1 - \frac{3}{2\sqrt[3]{x}}\right)$, whence it follows that as $x \to \infty$, the function also tends to infinity at about the same rate the function $2x$ does. The graph has the shape presented in Fig. 3.13. ▶

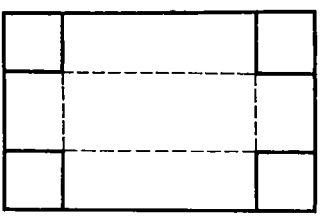


Fig. 3.14

Example 6. We make an open box from a rectangular piece of cardboard with sides 16 cm and 10 cm long by cutting out equal squares at the corners (Fig. 3.14) and bending along dashed lines. Try to make the volume of the box as large as possible.

Solution. Let x be the length (in cm) of a side of the square being cut out. Then the sides of the base of the box will be $16 - 2x$ and $10 - 2x$ long (in cm) and the height will be x (cm). Consequently, the volume of the box is

$$y = (16 - 2x)(10 - 2x)x = 4(x^3 - 13x^2 + 40x)\,(\text{cm}^3).$$

The function y is defined on the interval $(0, 5)$. We must find its absolute maximum on this interval. We find the derivative and equate it to zero. We get

$$3x^2 - 26x + 40 = 0, \quad x = \frac{13 \pm \sqrt{169 - 120}}{3},$$

$$x_1 = \frac{20}{3}, \quad x_2 = 2.$$

The first root $x_1 = 20/3$ does not lie on the interval $(0, 5)$. One "doubtful" point remains on this interval, the point $x_2 = 2$, and it yields the required maximum $y_{\max} = 72$ since at the endpoints of the interval the function vanishes and within the interval it is positive.

◀ 3.4. A Simple Example and Some Inferences from Its Consideration

Let us consider the function $z = \frac{\ln x}{x}$. It is defined on the interval $(0, +\infty)$. For $0 < x < 1$, its values are negative, and $z \to -\infty$ as $x \to 0$. To carry out the investigation, we calculate the derivative

$$z' = \frac{x \cdot \frac{1}{x} - \ln x}{x^2} = \frac{1 - \ln x}{x^2}$$

The denominator z' is positive. The numerator is positive for $\ln x < 1$, i.e. for $x < e$, and negative for $\ln x > 1$, i.e. for $x > e$. Consequently, on the interval $(0, e]$ the function z increases from $-\infty$ to the maximum $1/e$ for $x = e$ and vanishes for $x = 1$. On the interval $[e, +\infty)$ the function z decreases remaining positive. Therefore, on the interval $[1, +\infty)$ the function $\frac{\ln x}{x}$ is bounded by the number $1/e$ from above and by zero from below. To find out how the function in question behaves as $x \to \infty$, we shall go in a roundabout way, which will also yield some other interesting results.

We shall consider a function $w = \frac{\ln x}{x^a}$, where a is a positive number which may be very small. The behaviour of the function w resembles that of the function z. Its values are negative for $x < 1$ and $w \to -\infty$ as $x \to 0$. For $x > 1$ its values are positive. Furthermore,

$$w' = \frac{x^a \cdot \frac{1}{x} - \ln x \cdot a \cdot x^{a-1}}{x^{2a}} = \frac{1 - a \ln x}{x^{a+1}}.$$

The denominator of the fraction for w' is positive. The numerator is positive for $1 - a \ln x > 0$, i.e. for $x < e^{1/a}$, and becomes negative for $x > e^{1/a}$. For $x = e^{1/a}$ the function w attains its maximum equal to $\frac{\ln e^{1/a}}{(e^{1/a})^a} = \frac{1}{ae}$. Therefore the function w is bounded by the number $\frac{1}{ae}$ from above and, for $1 \leqslant x < \infty$, is bounded by zero from below.

From this follows the validity of the following interesting theorem.

Theorem. *For any* $a > 0$ *there holds a limiting relation* $\frac{\ln x}{x^a} \to 0$ *as* $x \to \infty$.

Proof. Indeed, $\frac{\ln x}{x^a} = \frac{\ln x}{x^{a/2}} \cdot \frac{1}{x^{a/2}}$. The first factor is bounded by virtue of what was said above. The second factor tends to zero since $x^{a/2} \to \infty$ as $x \to \infty$.

In particular, $z = \frac{\ln x}{x} \to 0$ as $x \to \infty$ so that the graph of this function goes down after the point $x = e$, $z = 1/e$

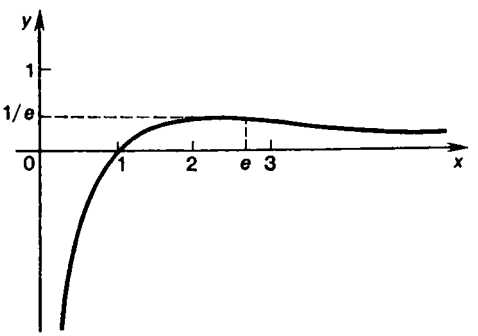


Fig. 3.15

and tends to the abscissa axis approaching it arbitrarily closely (Fig. 3.15).

Here is one more general remark. We say that an infinitely increasing function v tends to infinity much faster than the function w if $v/w \to \infty$, or, what is the same, as $w/v \to 0$. As $x \to \infty$, the function x^a tends to infinity if $a > 0$. Furthermore, $x^a \to \infty$ much faster than x^b as soon as $a > b > 0$ since $x^a/x^b = x^{a-b} \to \infty$ $(x \to \infty)$.

It immediately follows from the theorem that the power function x^a with any, even very small, positive exponent tends to infinity, as $x \to \infty$, much faster than a logarithmic function. Indeed, $\frac{\ln x}{x^a} \to 0$ as $x \to \infty$ and, hence, x^a tends to infinity much faster than $\ln x$. And what is more, x^a tends to infinity much faster than the function $(\ln x)^n$

for any, even large, exponent n since

$$\frac{(\ln x)^n}{x^a} = \left(\frac{\ln x}{x^{a/n}}\right)^n \to 0, \quad x \to \infty,$$

because

$$\frac{\ln x}{x^{a/n}} \to 0, \quad x \to \infty.$$

Let us return to the graph of the function $z = \frac{\ln x}{x}$ and apply the result of its investigation to the following rather difficult problem: we have to find the shape of the graph of the relation $x^y = y^x$ assuming, of course, that $x > 0$ and $y > 0$. It is clear that this graph includes a half-line $y = x$, $x > 0$, which coincides with the bisector of the first quadrant with the origin excluded. However, besides the points on this half-line, the graph has other points, say, (2, 4), since $2^4 = 4^2$.

We take logarithms of the equation and get $y \ln x = x \ln y$, whence it follows that $\frac{\ln x}{x} = \frac{\ln y}{y}$. Thus x and y must be the abscissas of the points on the graph of $z = \frac{\ln x}{x}$ which have the same ordinates. It can be seen from Fig. 3.15 that for $0 \leqslant x \leqslant 1$ there is only one possibility, $y = x$. When x slowly recedes from the point $x = 1$ to the right, another point with the same ordinate comes "from infinity" so that there is now one more possibility in addition to $y = x$. We designate the abscissa of this second point as y_2. As x increases from 1 to e (the point of maximum), y_2 varies from $+\infty$ to e, decreasing all the time. When x passes through the point of maximum, y_2 and x change parts so that $y_2 \to 1$ as $x \to \infty$. Thus the graph of the relation $x^y = y^x$ has the shape shown in Fig. 3.16. The point $x = 0$ is excluded. The point (2, 4) lies on the curvilinear portion of the graph. The infinite branches of the graph approach the straight lines $y = 1$ and $x = 1$ indefinitely. ▶

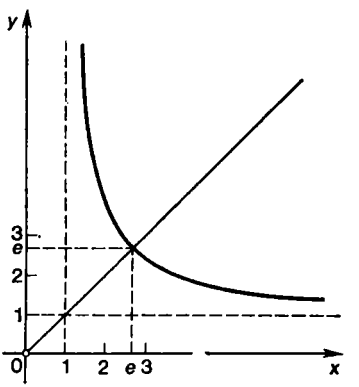


Fig. 3.16

3.5. Higher-Order Derivatives

In addition to considering the derivative $f'(x)$ of the function $f(x)$, it often becomes necessary to consider the derivative $f''(x)$ of the function $f'(x)$. It is known as the *second derivative* of the function $f(x)$. We know that the derivative is the rate of variation of a function. Therefore the second derivative is the rate of variation of the rate of variation of a function, or, using the term from mechanics, we can say that the second derivative is the acceleration of variation of a function.

Furthermore, the derivative of the second derivative is known as the *third derivative* or the *third-order derivative*, the derivative of the third derivative is the *fourth-order derivative* and so on. The derivative of order n of the function $f(x)$ is designated as $f^{(n)}(x)$ or $(f(x))^{(n)}$.

It is easy to see that $(f^{(k)}(x))^{(m)} = f^{(k+m)}(x)$. Indeed, every differentiation increases the order of the derivative by unity, and upon an m-fold differentiation, the order of the derivative increases by m so that the derivative of order k becomes a derivative of order $k + m$.

Here are some examples.

Example 1. Find the derivatives of all orders of the function x^4.

Solution. $(x^4)' = 4x^3$, $(x^4)'' = 12x^2$, $(x^4)''' = 24x$, $(x^4)^{(4)} = 24$, $(x^4)^{(n)} = 0$ for $n \geqslant 5$.

Example 2. Find the kth-order derivative of the function $f(x) = (x - a)^n$.

Solution. $f^{(k)}(x) = n(n-1) \ldots (n-k+1)(x-a)^{n-k}$ for $k \leqslant n$, $f^{(k)}(x) = 0$ for $k > n$.

Example 3. Find the nth-order derivative of the function $f(x) = e^{kx}$.

Solution. $f'(x) = ke^{kx}$, $f''(x) = k^2e^{kx}$, . . ., $f^{(n)}(x) = k^ne^{kx}$.

3.6. The Binomial Theorem

We know from algebra the formulas

$$(a + x)^2 = a^2 + 2ax + x^2,$$
$$(a + x)^3 = a^3 + 3a^2x + 3ax^2 + x^3.$$

It is easy to see that

$$(a + x)^4 = a^4 + 4a^3x + 6a^2x^2 + 4ax^3 + x^4,$$

$$(a + x)^5 = a^5 + 5a^4x + 10a^3x^2 + 10a^2x^3 + 5ax^4 + x^5$$

and so on. Let us derive a general formula for raising the sum of two terms into a power with any natural exponent.

It is clear that $(a + x)^n$ can be represented as a polynomial of x of degree n with coefficients expressed in terms of a. We expand this polynomial in increasing powers of x:

$$(a+x)^n = A_0 + A_1x + A_2x^2 + A_3x^3 + \ldots + A_kx^k + \ldots + A_{n-1}x^{n-1} + A_nx^n.$$

We calculate the derivatives up to the order n inclusive on both sides of this relation:

$$n(a+x)^{n-1} = A_1 + 2A_2x + 3A_3x^2 + \ldots + kA_kx^{k-1} + \ldots + (n-1)A_{n-1}x^{n-2} + nA_nx^{n-1},$$

$$n(n-1)(a+x)^{n-2} = 2A_2 + 3\cdot 2\cdot A_3x + \ldots + k(k-1)A_kx^{k-2} + \ldots + (n-1)(n-2)A_{n-1}x^{n-3} + n(n-1)A_nx^{n-2},$$

. . . . . .

$$n(n-1)\ldots(n-k+1)(a+x)^{n-k} = k(k-1)\ldots 2A_k + \ldots + n(n-1)\ldots(n-k+1)A_nx^{n-k},$$

. . . . . .

$$n(n-1)\ldots 2(a+x) = (n-1)(n-2)\ldots 2A_{n-1} + n(n-1)\ldots 2A_nx,$$

$$n(n-1)\ldots 2 = n(n-1)\ldots 2A_n.$$

Now we set $x = 0$ in all these relations and obtain

$$\begin{aligned}
a^n &= A_0,\\
na^{n-1} &= A_1,\\
n(n-1)\,a^{n-2} &= 2A_2,\\
&\dots\dots\dots\dots\\
n(n-1)\dots(n-k+1)\,a^{n-k} &= k(k-1)\dots 2A_k,\\
&\dots\dots\dots\dots\dots\dots\dots\dots\\
n(n-1)\dots 2a &= (n-1)(n-2)\dots 2A_{n-1},\\
n(n-1)\dots 2 &= n(n-1)\,2A_n,
\end{aligned}$$

whence it follows that

$$\begin{aligned}
A_0 &= a^n,\\
A_1 &= na^{n-1},\\
A_2 &= \frac{n(n-1)}{2}\,a^{n-2},\\
&\dots\dots\dots\dots\\
A_k &= \frac{n(n-1)\dots(n-k+1)}{2.3\dots k}\,a^{n-k},\\
&\dots\dots\dots\dots\dots\dots\\
A_{n-1} &= \frac{n(n-1)\dots 2}{2\dots(n-1)}\,a,\\
A_n &= \frac{n(n-1)\dots 2}{2.3\dots n}.
\end{aligned}$$

We transform somewhat the formula for the "general" coefficient A_k as follows. We introduce into consideration the product of all natural numbers from 1 to a certain m:

$$1\cdot 2\cdot\dots\cdot m.$$

This product is known as a *factorial* of the number m and is designated as $m!$. We multiply the numerator and denominator of the coefficient A_k by $(n-k)(n-k-1)\dots 2\cdot 1 = (n-k)!$ and write a factor 1 in the denominator. We obtain

$$\begin{aligned}
A_k &= \frac{n(n-1)\dots(n-k+1)(n-k)(n-k-1)\dots 2\cdot 1}{1.2.3\dots k\cdot(n-k)!}\,a^{n-k}\\
&= \frac{n!}{k!\,(n-k)!}\,a^{n-k}.
\end{aligned}$$

This is true for all k from 1 to $n-1$ and remains true for $k=0$ and $k=n$ if we assume that $0!=1$. The number $\frac{n!}{k!(n-k)!}$ is a *binomial coefficient* and is designated as C_k^n or $\binom{n}{k}$.

Thus

$$(a+x)^n = a^n + \binom{n}{1} a^{n-1}x + \binom{n}{2} a^{n-2}x^2 + \ldots + \binom{n}{k} a^{n-k}x^k + \ldots + \binom{n}{n-1} ax^{n-1} + x^n.$$

It is clear that $\binom{n}{k} = \binom{n}{n-k}$, and so the derived formula (known as the *binomial formula*) is symmetric with respect to the letters x and a.

Example 1. Remove the brackets in the expression $(a+2x)^8$.

Solution.

$$(a+2x)^8 = a^8 + 16a^7x + 112a^6x^2 + 448a^5x^3 + 1120a^4x^4 + 1792a^3x^5 + 1792a^2x^6 + 1024ax^7 + 256x^8.$$

Example 2. Calculate the number $(1.01)^{10}$ with an accuracy to within 10^{-5}.

Solution.

$$(1.01)^{10} \cong 1 + 10\cdot 10^{-2} + 45\cdot 10^{-4} + 120\cdot 10^{-6} = 1.10462.$$

Example 3. Compare the coefficients in x^{2m} in the identity

$$(1+x)^n (1-x)^n = (1-x^2)^n.$$

Solution.

$$\binom{n}{0}\binom{n}{2m} - \binom{n}{1}\binom{n}{2m-1} + \binom{n}{2}\binom{n}{2m-2} - \ldots + \binom{n}{2m}\binom{n}{0} = (-1)^m \binom{n}{m}.$$

3.7. Investigation of Functions with the Use of Higher-Order Derivatives

The first derivative $f'(x)$ of the function $f(x)$ has a clear geometrical meaning. It is the slope of a tangent, i.e. is equal to the tangent of the angle of inclination of the tangent to the abscissa axis. The second derivative is the rate

of variation of the slope of tangent. The positiveness of the second derivative on an interval means that the angle that the tangent makes with the abscissa axis increases with x. In terms of geometry this means that the graph is convex downwards. Now if the second derivative is negative on an interval, then its graph is convex upwards. In Fig. 3.17 the interval of definition of the function is divided into

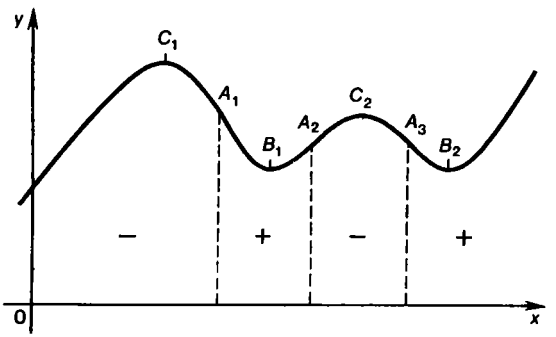


Fig. 3.17

parts, on each of which the second derivative retains sign (the sign is indicated in the figure). The points at which the graph changes the direction of convexity are *points of inflection* (points A_1, A_2, A_3 in Fig. 3.17). Passing through a point of inflection, the second derivative changes sign.

We can see that if at some point the first derivative is zero and the second derivative is positive (points B_1 and B_2 in Fig. 3.17), then the function has a minimum at that point since at such a point the tangent to the graph is horizontal and the graph is convex downwards. Correspondingly, if the first derivative is zero at a point and the second derivative is negative, then there is a maximum at that point (points C_1 and C_2 in Fig. 3.17).

◀ We shall now formulate and prove a theorem which extensively generalizes our arguments.

Theorem. *Assume that the function* $f(x)$, *defined on the interval* $[a, b]$, *has derivatives* $f'(x), \ldots, f^{(n-1)}(x), f^{(n)}(x)$, *and the relation* $f'(c) = \ldots = f^{(n-1)}(c) = 0$ *holds true at a point* $C \in [a, b]$.

Then, if $f^{(n)}(x) > 0$ *for all* $x \in [a, b]$, *then for an even* n *the function* $f(x)$ *has a minimum for* $x = c$, *and if* $n \geqslant 3$ *is odd, then the function* $f(x)$ *increases on* $[a, b]$ *and for it* $x = c$ *is a point of inflection. Correspondingly, if* $f^{(n)}(x) < 0$ *for all* $x \in [a, b]$, *then for an even* n *the function* $f(x)$ *has*

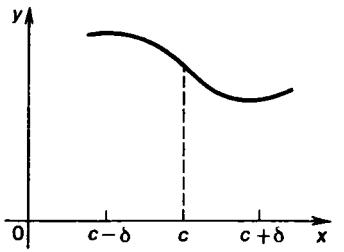


Fig. 3.18

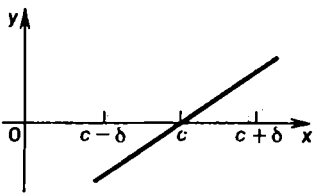


Fig. 3.19

a maximum at the point $x = c$ *and for an odd* $n \geqslant 3$ *the function* $f(x)$ *decreases on* $[a, b]$ *and for it* $x = c$ *is a point of inflection.*

Proof. Assume that the hypothesis is satisfied for $f^{(n)}(x) > 0$ (Fig. 3.18). Then $f^{(n-1)}(x)$ increases on the interval $[a, b]$ and so the relation

$$f^{(n-1)}(x) < f^{(n-1)}(c) = 0$$

holds true for $x < c$ and the relation

$$0 = f^{(n-1)}(c) < f^{(n-1)}(x)$$

holds true for $x > c$ (Fig. 3.19). Thus $f^{(n-1)}(x)$ is negative for $x < c$ and positive for $x > c$. Consequently, $f^{(n-2)}(x)$ decreases on the left of the point $x = c$ and increases on the right of this point. It vanishes for $x = c$. Therefore it assumes positive values on the left and on the right of the point $x = c$ and has a minimum for $x = c$ (Fig. 3.20). The function $f^{(n-3)}(x)$ increases on the left and on the right of the point $x = c$ so that it vanishes for $x = c$ and changes its values from negative to positive (Fig. 3.21). The function $f^{(n-4)}(x)$ decreases on the left of the point $x = c$ and increases on the right of it. Consequently, it has a minimum and is zero for $x = c$ and assumes positive values both on the left and on the right of c (Fig. 3.22). Presenting similar arguments, we find that $f^{(n-1)}(x)$, $f^{(n-3)}(x)$, $f^{(n-5)}(x)$, . . .

increase when x passes through the point $x = c$, and $f^{(n-2)}(x)$, $f^{(n-4)}(x)$, $f^{(n-6)}(x)$, . . . have a minimum for $x = c$. For an even n we reach the initial function $f(x)$

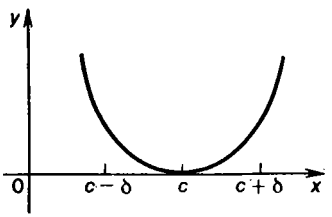


Fig. 3.20

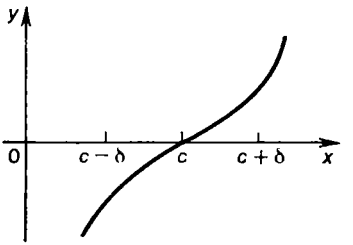


Fig. 3.21

after an even number of steps and infer that $f(x)$ has a minimum for $x = c$. For an odd n, we reach $f(x)$ after an odd number of steps and infer that $f(x)$ increases on the left of the point $x = c$ and keeps increasing on the right of it. Furthermore, $f''(x)$ also increases passing through the zero value, and, consequently, $f''(x)$ changes sign from minus to plus, and so the point c is a point of inflection for the function $f(x)$.

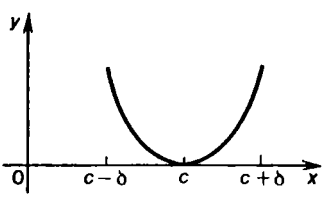


Fig. 3.22

The case $f^{(n)}(x) < 0$ can be considered in a similar way and so we omit it. We have proved the theorem.

Remark. If we assume in the hypothesis that $f^{(n)}(x) > 0$ not on the whole interval $[a, b]$ but only at the point $x = c$, then the inferences concerning the increase and decrease of the function $f(x)$ will remain valid but only, maybe, when we speak of the part $[c - \delta, c + \delta]$ of the interval $[a, b]$. Indeed, if $f^{(n)}(c) > 0$, then the function $f^{(n-1)}(x)$ increases at the point c and so there can be found $\delta > 0$ such that $f^{(n-1)}(x) < f^{(n-1)}(c) = 0$ for $c - \delta \leqslant x < c$ and $0 = f^{(n-1)}(c) < f^{(n-1)}(x)$ for $c < x \leqslant c + \delta$. All other arguments remain valid but only when we speak of the interval $[c - \delta, c + \delta]$.

Example 1. Using higher-order derivatives, test the function $f(x) = x^5 - 2x^4 + x^3 + 2$ for maximum and minimum.

Solution. $f'(x) = 5x^4 - 8x^3 + 3x^2 = x^2(5x^2 - 8x + 3)$. The roots of the derivative are $x_1 = 0$, $x_2 = 3/5$, $x_3 = 1$. There are three "doubtful" points. The second derivative

$$f''(x) = 20x^3 - 24x^2 + 6x, \quad f''(0) = 0,$$

$$f''\left(\frac{3}{5}\right) = -\frac{18}{25} < 0, \quad f''(1) = 2 > 0.$$

We infer that $x = 3/5$ is a point of maximum and $x = 1$ is a point of minimum. Furthermore,

$$f'''(x) = 60x^2 - 48x + 6, \quad f'''(0) = 6 > 0.$$

Consequently, the point $x = 0$ is an inflection point on the increase. The graph has the shape shown in Fig. 3.23.

Remark. We could have certainly solved this example by using only the test for signs of $f'(x)$. The procedure is very simple when the function is factorized:

$$f'(x) = x^2(5x - 3)(x - 1).$$

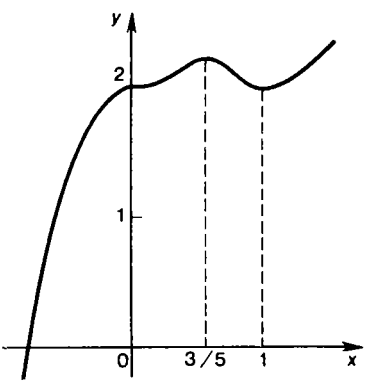


Fig. 3.23

Example 2. Prove that if the function $f(x)$ has a continuous derivative, then there is at least one inflection point between the neighbouring points of local maximum and local minimum.

Solution. At the point of maximum the convexity is directed upwards and at the point of minimum downwards. The direction of convexity must change upon a passage through a certain point at least once.

Example 3. Find the inflection points of the graph of the function $y = \frac{x}{1+x^2}$.

Solution. $y' = \frac{1-x^2}{(1+x^2)^2}$, $y'' = \frac{2x(x^2-3)}{(1+x^2)^3}$.

The function y'' vanishes at the points $x = 0$, $x = -\sqrt{3}$ and $x = \sqrt{3}$ and changes sign when x passes through every one of these points. Consequently, points with abscissas $-\sqrt{3}$, 0 and $\sqrt{3}$ are inflection points. ▶

3.8. The Order of Smallness of a Function in the Neighbourhood of a Point at Which the Function Vanishes and the Proximity of Functions

Consider the functions $x - a$, $(x - a)^2$, $(x - a)^3$, ..., $(x - a)^n$, ... They all vanish for $x = a$ and tend to zero as $x \to a$. And each of these functions tends to zero much faster than the preceding one. Indeed, $\frac{(x-a)^m}{(x-a)^k} = (x - a)^{m-k} \to 0$ for $m > k$ and as $x \to a$. The following numerical examples show this especially vividly: if $x - a = 1/2$, then $(x - a)^2 = 1/4$, i.e. the difference is half as large, $(x - a)^3 = 1/8$, half as large again and so on. And for $x - a = 1/10$, every successive power $(x - a)^{k+1}$ is ten times smaller than the preceding power $(x - a)^k$. For $x - a$ still smaller, the decrease in $(x - a)^k$, with an increase in k, becomes faster still.

If the continuous function $f(x)$ vanishes for $x = a$, then it tends to zero as $x \to a$. It is natural to compare $f(x)$ with the powers $(x - a)^n$ in the sense of the rate of tending to zero. Namely, the *order of smallness* of $f(x)$ for $x \to a$ is considered to be equal to n if there is a nonzero limit of $f(x)/(x - a)^n$ as $x \to a$. Thus, $f(x) = x^3 + x^4 + x^5$ has an order of smallness 3 as $x \to 0$, $f(x)/x^3 = 1 + x + x^2 \to 1$ as $x \to 0$. The function $\sqrt{x} - 2$ has an order of smallness 1 as $x \to 4$ since $\frac{\sqrt{x}-2}{x-4} = \frac{1}{\sqrt{x}+2} \to \frac{1}{4}$ as $x \to 4$ and so on.

We can also introduce fractional orders of smallness for some functions, assuming that the function $|x - a|^\alpha$ has an order of smallness α for $\alpha > 0$. Thus the function $x^{3/2}$ has an order of smallness 3/2 in the neighbourhood of $x = 0$, the function $\ln(1 + \sqrt[3]{x})$ has an order of smallness 1/3 in the neighbourhood of $x = 0$ since

$$\lim_{x \to 0} \frac{\ln(1+\sqrt[3]{x})}{\sqrt[3]{x}} = 1 \neq 0.$$

There are functions tending to zero as $x \to a$ to which we cannot assign a definite order of smallness. It is expedient to

introduce other definitions to characterize the order of smallness of such functions. Namely, we say that the order of smallness of the function $f(x)$ in the neighbourhood of the point $x = a$ is not lower than n if the function $\frac{f(x)}{(x-a)^n}$ is bounded in the neighbourhood of the point $x = a$. It is almost evident that if the order of smallness of $f(x)$ is equal to or higher than n in the sense of the previous definition, it is not lower than n in the sense of the last definition either. Indeed, if the order of smallness of $f(x)$ is $m \geqslant n$, then $\frac{f(x)}{(x-a)^n} = \frac{f(x)}{(x-a)^m}(x-a)^{m-n}$ tends to zero for $m > n$ or to a finite limit if $m = n$ and, consequently, $\frac{f(x)}{(x-a)^n}$ is bounded in the neighbourhood of the point $x = a$.

The *proximity of two functions* in the neighbourhood of a given point is the order of smallness of their difference. Thus the functions $y = x^2$ and $z = 2x - 1$ have a proximity of order 2 in the neighbourhood of the point $x = 1$ since $y - z = x^2 - 2x + 1 = (x-1)^2$. Here we have taken as z a linear function whose graph is a tangent to the parabola $y = x^2$ at the point $x = 1$. We can prove, and we shall do it in what follows, that the proximity of a function to a linear function, which defines its tangent, is of order 2 if the second derivative exists at the point of tangency and is different from zero.

It is natural that for the given function we can choose not only a linear function which is close to it but also a quadratic trinomial and also polynomials of a higher degree which ensure still higher proximity. For the time being, we shall not discuss this problem theoretically but shall consider some examples.

Example 1. For the function $y = x^4$ choose a quadratic trinomial which would have a proximity of order 3 in the neighbourhood of the point $x = 1$.

Solution. $x^4 = (1 + (x-1))^4 = 1 + 4(x-1) + 6(x-1)^2 + 4(x-1)^3 + (x-1)^4$. It is clear that the required trinomial is $z = 1 + 4(x-1) + 6(x-1)^2 = 6x^2 - 8x + 3$ since $y - z = 4(x-1)^3 + (x-1)^4 = (x-1)^3(4 + x - 1)$ has a third order of smallness.

Example 2. For the function $y = \sqrt{1+x}$ choose a quadratic trinomial $z = a + bx + cx^2$ which would have a

proximity of order 3 to the function y in the neighbourhood of the point $x = 0$.

Solution. We must take $a = 1$ for the relation $y|_{x=0} = z|_{x=0}$. Furthermore,

$$y - z = \sqrt{1+x} - (1 + bx + cx^2) = \frac{1 + x - (1 + bx + cx^2)^2}{\sqrt{1+x} + (1 + bx + cx^2)}$$

$$= \frac{1 + x - 1 - 2bx - (b^2 + 2c)\, x^2 - 2bcx^3 - c^2x^4}{\sqrt{1+x} + 1 + bx + cx^2}$$

$$= \frac{(1 - 2b)\, x - (b^2 + 2c)\, x^2 - 2bcx^3 - c^2x^4}{\sqrt{1+x} + 1 + bx + cx^2}.$$

We take $2b = 1$ so that $b = 1/2$ and let $b^2 + 2c = 0$ so that $c = -1/8$. Then

$$y - z = \frac{\frac{1}{8}x^3 - \frac{1}{64}x^4}{\sqrt{1+x} + 1 + \frac{1}{2}x - \frac{1}{8}x^2},$$

$$\frac{y-z}{x^3} = \frac{\frac{1}{8} - \frac{1}{64}x}{\sqrt{1+x} + 1 + \frac{1}{2}x - \frac{1}{8}x^2} \to \frac{1}{16}$$

as $x \to 0$. Consequently, $y - z$ has a third order of smallness in the neighbourhood of the point $x = 0$, and this is what we wanted to prove. ▶

◀ 3.9. The Relationship Between the Order of Smallness and the Order of the First Nonzero Derivative

We can use another approach to the order of smallness of a function in the neighbourhood of its root. If a function vanishes for $x = a$ but its derivative does not, then the graph of the function cuts the abscissa axis at a nonzero angle. If the derivative is also zero, then the graph of the function touches the abscissa axis at one point and the whole graph comes closer to the abscissa axis than in the first case. Now if, in addition, the second derivative is zero, then, farther from the point $x = a$, not only the rate of variation of the function is zero but also the acceleration so that the graph adheres to the abscissa axis still closer. This effect becomes

stronger with the increase in the number of derivatives which vanish for $x = a$.

For the function $(x - a)^n$ whose order of smallness is n in the neighbourhood of $x = a$ we have a case when all its derivatives from the first to the $(n - 1)$th vanish for $x = a$ and the nth derivative, equal to $n!$, is nonzero.

The following theorem is valid.

Theorem. *If* $f(a) = f'(a) = \ldots = f^{(n-1)}(a) = 0$ *and* $f^{(n)}(x)$ *exists in a neighbourhood* $(a - \delta, a + \delta)$ *of the point* a *and* $m < f^{(n)}(x) < M$ *at all points of this neighbourhood, then all the values of* $f(x)$ *lie between* $\frac{m(x-a)^n}{n!}$ *and* $\frac{M(x-a)^n}{n!}$.

Proof. First assume that n is even and consider functions

$$g(x) = f(x) - m\frac{(x-a)^n}{n!}, \quad h(x) = f(x) - \frac{M(x-a)^n}{n!}.$$

It is easy to see that $g(a) = g'(a) = \ldots = g^{(n-1)}(a) = 0$ and $h(a) = h'(a) = \ldots = h^{(n-1)}(a) = 0$ since both terms the functions $g(x)$ and $h(x)$ are composed of possess this property. Furthermore,

$$g^{(n)}(x) = f^{(n)}(x) - m\frac{n!}{n!} = f^{(n)}(x) - m > 0$$

for all x from the neighbourhood being considered. Consequently, by virtue of the theorem from 3.7, the function $g(x)$ has a minimum at the point $x = a$ for the zero value, and so $g(x) > 0$ for $x \neq a$ and $x \in [a - \delta, a + \delta]$ and, consequently, $f(x) > m(x - a)^n/n!$. Similarly, $h^{(n)}(x) = f^{(n)}(x) - M < 0$, and so $h(x)$ has a maximum for the zero value at the point $x = a$, and so $h(x) < 0$ for $x \neq a$, $x \in [a - \delta, a + \delta]$, i.e. $f(x) < M(x - a)^n/n!$.

Thus, for an even n we have proved that for $x \neq a$, $x \in [a - \delta, a + \delta]$, there hold inequalities

$$\frac{m(x-a)^n}{n!} < f(x) < \frac{M(x-a)^n}{n!}.$$

Let now n be odd. As before, for functions $g(x)$ and $h(x)$ there hold equalities $g(a) = g'(a) = \ldots = g^{(n-1)}(a) = 0$, $h(a) = h'(a) = \ldots = h^{(n-1)}(a) = 0$ and inequalities $g^{(n)}(x) = f^{(n)}(x) - m > 0$ and $h^{(n)}(x) = f^{(n)}(x) - M < 0$. On the basis of the theorem from 3.7 we infer that $g(x)$ increases, passing through zero for $x = a$, and $h(x)$ decreases,

and so $g(x) < 0$ for $a - \delta \leqslant x < a$, $g(x) > 0$ for $a < x \leqslant a + \delta$, $h(x) > 0$ for $a - \delta \leqslant x < a$ and $h(x) < 0$ for $a < x \leqslant a + \delta$. Thus, for $a < x \leqslant a + \delta$ there hold inequalities $\frac{m(x-a)^n}{n!} < f(x) < \frac{M(x-a)^n}{n!}$, and for $a - \delta \leqslant x < a$ there hold inequalities $\frac{M(x-a)^n}{n!} < f(x) < \frac{m(x-a)^n}{n!}$.

We have proved the theorem.

Remark 1. In the hypothesis, we can include equality signs into the estimates for $f^{(n)}(x)$, i.e. assume that $m \leqslant f^{(n)}(x) \leqslant M$ for $x \in [a - \delta, a + \delta]$. These inequalities will be satisfied if we take the least and the greatest value of $f^{(n)}(x)$ on $[a - \delta, a + \delta]$ as m and M. Then in the conclusion of the theorem we must also include equality signs in the inequalities

$$\frac{m(x-a)^n}{n!} < f(x) < \frac{M(x-a)^n}{n!},$$

$$\frac{M(x-a)^n}{n!} < f(x) < \frac{m(x-a)^n}{n!}$$

(the last one for an odd n and $x < a$). Indeed, if $m \leqslant f^{(n)}(x) \leqslant M$, then $m - \alpha < f^{(n)}(x) < M + \alpha$ for any $\alpha > 0$ and, consequently (for an even or odd n but for $x > a$),

$$\frac{(m-\alpha)(x-a)^n}{n!} < f(x) < \frac{(M+\alpha)(x-a)^n}{n!},$$

and, since we can take α arbitrarily small,

$$\frac{m(x-a)^n}{n!} \leqslant f(x) \leqslant \frac{M(x-a)^n}{n!}.$$

Correspondingly, for an odd n and $x < a$ we have

$$\frac{M(x-a)^n}{n!} \leqslant f(x) \leqslant \frac{m(x-a)^n}{n!}.$$

Remark 2. The inequalities we have proved are equivalent to the inequalities $\frac{m}{n!} \leqslant \frac{f(x)}{(x-a)^n} \leqslant \frac{M}{n!}$ which hold true for an even as well as for an odd n. To verify this, it is sufficient to divide the three parts of the inequalities we have

proved by $(x-a)^n$. Both for an even and for an odd n but for $x > a$ the number $(x-a)^n$ is positive and the inequality signs must be retained. For an odd n and $x < a$, the number $(x-a)^n$ is negative and after division we must change the inequality signs to the opposite, which in this case as well leads to the inequalities $\frac{m}{n!} \leqslant \frac{f(x)}{(x-a)^n} \leqslant \frac{M}{n!}$. We set $\frac{f(x)}{(x-a)^n/n!} = T(x)$ and obtain $m \leqslant T(x) \leqslant M$ and $f(x) = \frac{(x-a)^n}{n!} T(x)$.

Corollary. *If* $f(a) = f'(a) = \ldots = f^{(n-1)}(a) = 0$ *and* $f^{(n)}(x)$ *is continuous in the neighbourhood of the point* a, *then there is a limit of* $\frac{f(x)}{(x-a)^n}$ *as* $x \to a$ *and this limit is equal to* $\frac{f^{(n)}(a)}{n!}$.

Proof. Indeed, if we assume that $\delta \to 0$, i.e. take an infinitely decreasing neighbourhood of the point a, then the least m and the greatest M value of $f^{(n)}(x)$ on $[a-\delta, a+\delta]$ tend to $f^{(n)}(a)$, so that

$$\lim_{x \to a} \frac{f(x)}{(x-a)^n} = \lim_{x \to a} \frac{m}{n!} = \lim_{x \to a} \frac{M}{n!} = \frac{f^{(n)}(a)}{n!}.$$

Therefore, if $f(a) = f'(a) = \ldots = f^{(n-1)}(a) = 0$, $f^{(n)}(x)$ exists and is continuous in the neighbourhood of the point $x = a$ and $f^{(n)}(a) \neq 0$, then the order of smallness of the function $f(x)$ in the neighbourhood of the point a is exactly equal to n.

Thus, if the function $f(x)$ has derivatives of all orders in the neighbourhood of the point $x = a$ and $f(a) = 0$, then the order of smallness of $f(x)$ in the neighbourhood of the point $x = a$ is exactly equal to the lowest order of the derivative which does not vanish at the point a.

In these natural assumptions, the order of smallness of the function $f(x)$ is an integer. Consequently, functions whose order of smallness is fractional must not have continuous derivatives of a sufficiently high order in the neighbourhood of the point $x = a$. Thus the function $x^{3/2}$ whose order of smallness is 3/2 in the vicinity of $x = 0$, has a second derivative $\frac{3}{4}x^{-1/2}$ which has a discontinuity for $x = 0$.

Example 1. Prove that the order of smallness of the function

$$f(x)=\sqrt{x}-\frac{5}{16}-\frac{15}{16}x+\frac{5}{16}x^2-\frac{1}{16}x^3$$

in the vicinity of the point $x=1$ is 4 and find $\lim\limits_{x\to 1}\frac{f(x)}{(x-1)^4}$.

Solution.

$$f(1)=0,\quad f'(x)=\frac{1}{2\sqrt{x}}-\frac{15}{16}+\frac{10}{16}x-\frac{3}{16}x^2,\quad f'(1)=0,$$

$$f''(x)=-\frac{1}{4\sqrt{x^3}}+\frac{5}{8}-\frac{3}{8}x,\quad f''(1)=0,$$

$$f'''(x)=\frac{3}{8\sqrt{x^5}}-\frac{3}{8},\quad f'''(1)=0,$$

$$f^{(4)}(x)=-\frac{15}{16\sqrt{x^7}},\quad f^{(4)}(1)=-\frac{15}{16}\neq 0.$$

Thus the order of smallness of $f(x)$ in the neighbourhood of the point $x=1$ is 4 indeed. Furthermore,

$$\lim_{x\to 1}\frac{f(x)}{(x-1)^4}=\frac{f^{(4)}(1)}{4!}=-\frac{5}{128}.$$

Example 2. Prove that the linear function $l(x)=f(a)+f'(a)(x-a)$, whose graph is a tangent to the graph of the function $f(x)$ at the point $x=a$, has a proximity of order 2 to the function $f(x)$ if $f''(a)\neq 0$.

Proof. $g(x)=l(x)-f(x)=f(x)-f(a)-f'(a)(x-a)$. Therefore, $g(a)=0$. Furthermore, $g'(x)=f'(x)-f'(a)$ and $g'(a)=0$. Finally, $g''(x)=f''(x)$ and $g''(a)=f''(a)\neq 0$. Consequently, the order of smallness of the function $g(x)$ is 2 in the neighbourhood of the point $x=a$, i.e. the proximity of $l(x)$ and $f(x)$ is also of order 2.

Example 3. Find a third-degree polynomial $P(x)$ that would have a proximity of order 4 to the function e^x in the neighbourhood of the point $x=0$.

Solution. Assume that

$$P(x)=a_0+a_1x+a_2x^2+a_3x^3,$$

$$g(x)=e^x-P(x)=e^x-a_0-a_1x-a_2x^2-a_3x^3.$$

We must choose coefficients such that the function $(g)\,x$ would have an order of smallness 4 in the neighbourhood of

$x = 0$. For this purpose, we must require that $g(0) = g'(0) = g''(0) = g'''(0) = 0$. But

$$g(0) = 1 - a_0,$$
$$g'(x) = e^x - a_1 - 2a_2x - 3a_3x^2, \quad g'(0) = 1 - a_1,$$
$$g''(x) = e^x - 2a_2 - 6a_3x, \quad g''(0) = 1 - 2a_2,$$
$$g'''(x) = e^x - 6a_3, \quad g'''(0) = 1 - 6a_3.$$

Finally, $g^{(4)}(x) = e^x$ and $g^{(4)}(0) = 1 \neq 0$. Equating $g(0)$, $g'(0)$, $g''(0)$ and $g'''(0)$ to zero, we find that $a_0 = 1$, $a_1 = 1$, $a_2 = 1/2$, $a_3 = 1/6$. Thus the required polynomial $P(x)$ is $1 + x + \frac{1}{2}x^2 + \frac{1}{6}x^3$. ▶

◀ 3.10. Formulas of Taylor and Maclaurin

For the function $f(x)$ which has $n + 1$ continuous derivatives in the neighbourhood of the point $x = a$ we can always find a polynomial $P_n(x)$ of degree n such that it would have a proximity to $f(x)$ not lower than $n + 1$ in the neighbourhood of the point $x = a$. Let us prove this.

Assume that $P_n(x) = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n$. We write $x = (x - a) + a$ and use the binomial theorem to transform the degrees of x. We represent each term a_kx^k as the sum of the degrees of the binomial $(x - a)$ with some coefficients. Collecting terms, we represent $P_n(x)$ in the form

$$P_n(x) = b_0 + b_1(x - a) + b_2(x - a)^2 + \ldots + b_n(x - a)^n.$$

It is convenient for us to seek $P_n(x)$ in precisely this form.

We set $r_n(x) = f(x) - P_n(x) = f(x) - b_0 - b_1(x - a) - b_2(x - a)^2 - \ldots - b_k(x - a)^k - \ldots - b_n(x - a)^n$. To achieve our purpose, we must require that $r_n(a) = r_n'(a) = \ldots = r_n^{(n)}(a) = 0$. We calculate the derivatives of the function $r_n(x)$:

$$r_n'(x) = f'(x) - b_1 - 2b_2(x - a) - \ldots - kb_k(x - a)^{k-1} - \ldots - nb_n(x - a)^{n-1},$$
$$r_n''(x) = f''(x) - 2!b_2 - 3 \cdot 2 \cdot b_3(x - a) - \ldots - k(k - 1)b_k(x - a)^{k-2} - \ldots - n(n - 1)b_n(x - a)^{n-2},$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$

$$r_n^{(k)}(x) = f^{(k)}(x) - k!\, b_k - \ldots - n(n-1) \ldots (n-k+1)\, b_n (x-a)^{n-k},$$

. . . . . .

$$r_n^{(n-1)}(x) = f^{(n-1)}(x) - (n-1)!\, b_{n-1} - n(n-1) \ldots 2b_n (x-a),$$

$$r_n^{(n)}(x) = f^{(n)}(x) - n!\, b_n.$$

Equating $r_n(a)$, $r_n'(a)$, ..., $r_n^{(n)}(a)$ to zero, we obtain

$$\begin{aligned}
f(a) - b_0 &= 0,\\
f'(a) - b_1 &= 0,\\
f''(a) - 2!\, b_2 &= 0,\\
&\ldots\ldots\ldots\ldots\\
f^{(k)}(a) - k!\, b_k &= 0,\\
&\ldots\ldots\ldots\ldots\\
f^{(n-1)}(a) - (n-1)!\, b_{n-1} &= 0,\\
f^{(n)}(a) - n!\, b_n &= 0,
\end{aligned}$$

whence $b_0 = f(a)$, $b_1 = f'(a)$, $b_2 = \frac{1}{2!} f''(a)$, ..., $b_k = \frac{1}{k!} f^{(k)}(a)$, ..., $b_{n-1} = \frac{1}{(n-1)!} f^{(n-1)}(a)$, $b_n = \frac{1}{n!} f^{(n)}(a)$.

With this choice of coefficients, the function $r_n(x)$ will have an order of smallness not lower than $n+1$ and, correspondingly, $P(x)$ will have a proximity to $f(x)$ not lower than $n+1$. Thus we have

$$f(x) = P_n(x) + r_n(x) = f(a) + f'(a)(x-a) + \frac{1}{2!} f''(a)(x-a)^2 + \ldots + \frac{1}{k!} f^{(k)}(a)(x-a)^k + \ldots + \frac{1}{n!} f^{(n)}(a)(x-a)^n + r_n(x),$$

where $r_n(x)$ is a function whose order of smallness in the neighbourhood of the point $x = a$ is not lower than $n+1$.

The formula obtained is known as *Taylor's formula*. The function $r_n(x)$ is the *remainder* in Taylor's formula. On the basis of the theorem from 3.7, we can estimate the remainder. Clearly, $r_n^{(n+1)}(x) = f^{(n+1)}(x)$ and, by construction, $r(a) = r'(a) = \ldots = r^{(n)}(a) = 0$. According to the the-

orem from 3.7, we have

$$r_n(x) = \frac{(x-a)^{n+1}}{(n+1)!} T_n(x),$$

where $m \leqslant T_n(x) \leqslant M$, the numbers m and M being the least and the greatest value of $f^{(n+1)}(x)$ in the neighbourhood of the point $x = a$ being considered. The remainder characterizes the error resulting from the replacement of the values of the function $f(x)$ by those of the polynomial $P_n(x)$.

In a particular case, for $a = 0$, the formula assumes a simpler form

$$f(x) = f(0) + f'(0)\,x + \frac{1}{2!} f''(0)\,x^2 + \ldots + \frac{1}{n!} f^{(n)}(0)\,x^n + r_n(x).$$

This formula is known as *Maclaurin's formula.*

Let us consider some examples.

Example 1. Expand the polynomial

$$P(x) = x^5 - 2x^4 + 3x^3 - x^2 + x - 1$$

in the powers of the binomial $(x - 1)$.

First we seek all derivatives of $P(x)$ up to the fifth order inclusive:

$$\begin{aligned} P'(x) &= 5x^4 - 8x^3 + 9x^2 - 2x + 1, \\ P''(x) &= 20x^3 - 24x^2 + 18x - 2, \\ P'''(x) &= 60x^2 - 48x + 18, \\ P^{(4)}(x) &= 120x - 48, \\ P^{(5)}(x) &= 120. \end{aligned}$$

Then we calculate the values of the polynomial and its derivatives at the point $x = 1$:

$$P(1) = 1, \quad P'(1) = 10, \quad P''(1) = 12,$$
$$P'''(1) = 30, \quad P^{(4)}(1) = 72, \quad P^{(5)}(1) = 120.$$

From Taylor's formula we obtain

$$P(x) = 1 + 10(x-1) + \frac{12}{2!}(x-1)^2 + \frac{30}{3!}(x-1)^3 + \frac{72}{4!}(x-1)^4 + \frac{120}{5!}(x-1)^5,$$

The final form is

$$P(x) = 1 + 10(x-1) + 6(x-1)^2 + 5(x-1)^3 + 3(x-1)^4 + (x-1)^5.$$

We shall get the same result if we use the technique employed in the solution of Example 1 from 3.8.

Making use of this expansion, we can calculate the value of $P(x)$ for any value of x. But it is evident that for x which are far from unity, such a form of representation of a polynomial has no advantages as compared to the given one. However, the situation changes radically when we pose a problem of approximate calculation of $P(x)$ for x which are close to 1. Then the calculations become simpler and, in addition, the error can be estimated beforehand. For instance, let us calculate $P(1.1)$ with an accuracy of 0.001. The last two terms introduce a correction smaller than 0.001 and therefore

$$P(1.1) \cong 1 + 10 \cdot 0.1 + 6 \cdot 0.01 + 5 \cdot 0.001 = 2.065.$$

Example 2. Use Maclaurin's formula to expand the function $f(x) = e^x$.

Solution. Since $f^{(n)}(x) = e^x$ for any n, we have $f^{(n)}(0) = e^0 = 1$ and get, from Maclaurin's formula,

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + r_n(x).$$

Note that for $x = 1$ we get a formula for an approximate calculation of the number e:

$$e \cong 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots + \frac{1}{n!}.$$

The error of calculation is estimated as follows:

$$0 < r_n(1) < e\frac{1}{(n+1)!} < \frac{3}{(n+1)!}.$$

Let us calculate the value of the number e with an accuracy of 10^{-3}. To ensure this accuracy, it is sufficient to take $n = 6$ and then

$$0 < r_6 < \frac{3}{7!} = \frac{3}{5040} < 0.0006,$$

$$e \cong 1 + 1 + \frac{1}{1 \cdot 2} + \frac{1}{1 \cdot 2 \cdot 3} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} + \frac{1}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5 \cdot 6}$$

$$= 2 + \frac{360 + 120 + 30 + 6 + 1}{720} = 2\frac{517}{720} \cong 2.718. \blacktriangleright$$

3.11. General Concepts of the Theory of Approximation

When using mathematics in the solution of applied problems, we often have to deal with numbers given approximately. In some cases we do not know the exact value of the error in the given numbers but only have its estimate in absolute value. If the approximate value of x is x_0 and it is known that the error does not exceed α in absolute value, then the accepted notation is $x = x_0 \pm \alpha$. This is a conventional notation and must not be understood literally (x is equal to $x_0 + \alpha$ or $x_0 - \alpha$). Thus the notation $x = 1.324 \pm 5 \cdot 10^{-4}$ means that x is approximately equal to 1.324 and the error does not exceed five units in the fourth decimal place.

Sometimes it becomes necessary, for some reason, to discard some decimal digits in the notation of a number. This operation is known as *rounding off*. If the first discarded digit is smaller than 5 when we round off a number, then we leave the preceding digit without changes, if it is larger than 5 then we increase the preceding digit by unity. If the first discarded digit is 5, we can round off the number "downwards" (without changing the preceding digit) or "upwards" (increasing the preceding digit by unity). Thus, for the number e for which the rounded value with seven decimal digits is 2.7182818, the numbers 2.7, 2.72, 2.718, 2.7183, 2.71828, 2.718282 (with one, two, . . ., six decimal digits) are less precise approximations.

If x_0 is a rounded value of x taken with n decimal digits, then we can write

$$x = x_0 \pm 0.5 \cdot 10^{-n}.$$

The difference between the true value of a quantity and its approximate value is the *absolute error* of approximation. As we know by now, in approximate calculations the exact value of the absolute error is not usually known beforehand and only its *estimate* is known, i.e. the number which bounds its absolute value from above. If we use an approximate notation of a number in the form of a finite decimal fraction and know that all the digits we have written are correct, then the absolute error is estimated by five units of the first discarded decimal place. Thus, when we say that a number

is defined with six valid decimal digits, this means that the number $5 \cdot 10^{-7}$ is the estimate of the absolute error.

The *relative error* is the ratio of the value of the absolute error to the approximate value of the quantity. The accuracy of the approximate value of a quantity is usually defined by the estimate of the relative error. This estimate is often given in per cent or fractions of a per cent. Thus, if we know that the value of a quantity is 0.01232 with an accuracy of 0.1%, this means that the relative error does not exceed $0.001 = 10^{-3}$ and the absolute error does not exceed $10^{-3} \cdot 0.01232 = 0.00001232 \cong 0.00001$, i.e. the fifth decimal digit is doubtful and may differ from the true value by one unit.

The relative error possesses the following remarkable property.

Theorem 1. *The relative error of a quantity defined approximately does not change when this quantity is multiplied by an exact number.*

Proof. Indeed, assume that x is a quantity whose approximate value is x_0 and c is a number whose exact value is known. Then cx_0 is an approximate value of cx and the relative error of the quantity cx is $\frac{cx - cx_0}{cx_0} = \frac{x - x_0}{x_0}$, i.e. equal to the relative error for the quantity x.

The case $c = 10^n$, where n is an integer, positive or negative, is especially important for applications. Multiplication by 10^n does not change the number of *significant digits* in the notation (i.e. the number of digits between the first nonzero digit and the last reliable one, which may be followed by doubtful digits, which should in turn be either discarded or replaced by zeros). Thus the numbers 12345000 and 0.0012345, which have the same digits considered to be valid, can be used for the notation of approximate values of quantities with the same estimate of the relative error:

$$\frac{0.5}{12345} \cong 0.00004 = 0.004\%.$$

Computers often output results in floating point format, i.e. the computer points a zero, a decimal point, and the significant digits (*mantissa*) and then points the "exponent", i.e. the power to which the number 10 is raised which is then multiplied by the mantissa, i.e. the place is indicated

where the decimal point must be placed, or, for large numbers, how many zeros must be placed beside the mantissa. Of course, when we are accustomed to scientific notation with a mantissa and an exponent (or characteristic), this notation becomes even more useful than the ordinary one. Thus the notations 0.12345×10^6 and 0.12345×10^{-3} are as easy to grasp as 123 450 and 0.00012345.

A computational problem is considered to be *well-posed* (*well conditioned*) if the relative error of the result of computations exceeds but little the relative errors of the input data. One of the causes of a poor conditioning is the "disappearance of significant digits". This unpleasant phenomenon occurs, in particular, when we subtract two close numbers from one another (in explicit or implicit form). Thus, for instance, the numbers 0.1234 and 0.1222 with valid significant digits are given approximately, with the same error not exceeding $\frac{5 \cdot 10^{-5}}{0.12} \cong 4 \cdot 10^{-4}$, and their difference 0.0012 even has a doubtful fourth decimal digit, so that the relative error is estimated by the number $\frac{10^{-4}}{0.0012} = 6 \cdot 10^{-2}$, i.e. in the sense of the estimation of the relative error, the result is approximately 200 times worse than the input data.

When we solve a complicated problem in mathematics and choose a method of computation, we must try to avoid, if possible, the loss of significant digits.

3.12. Estimation of the Errors of the Results of Calculations Involving Approximate Numbers

To judge the accuracy of approximate calculations from the estimates of the errors, it is sufficient to have a rough approximate knowledge of these estimates. Sometimes it is sufficient to know only the exact order of the estimate (for instance, the absolute error is estimated by the phrase "the first three decimal digits are correct" and the phrase "the result is correct with an accuracy to within several tenths of a percent" estimates the relative error). We can regard the unknown exact magnitude of the absolute error as small increments to the approximate value and when we speak of using an approximately defined argument to calculate a function, we can replace the increment by the differential.

Theorem 1. *The absolute error in the calculation of the value of the function $y = f(x)$ of the approximately defined argument is approximately the product of the absolute error of the argument by the value of the derivative $f'(x)$ at the point in question.*

Proof. Indeed, assume that x_0 is an approximate value of the argument and x is the unknown exact value of the argument obviously close to x_0. We can assume the value $\Delta x = x - x_0$ of the absolute error of the argument to be small increments of the argument. The corresponding increment $\Delta y = f(x) - f(x_0)$ of the function, i.e. the absolute error for the value of the function, is approximately equal to the differential $dy = f'(x_0)\,\Delta x = f'(x_0)(x - x_0)$, and this is what we wished to prove.

It stands to reason that the theorem is true on the assumption that the absolute error of the argument is in fact sufficiently small for the given function.

It follows from Theorem 1 that the estimate of the absolute error of the value of the function does not exceed the estimate of the absolute error of the argument multiplied by the absolute value of the derivative.

The following formula holds true for the relative error:

$$\frac{\Delta y}{y_0} \approx \frac{f'(x_0)\,\Delta x}{f(x_0)} = \frac{x_0 \cdot f'(x_0)}{f(x_0)} \cdot \frac{\Delta x}{x_0}.$$

If the accuracy of calculations is judged by the increase in the relative error, then we have to consider the calculation of the function to be "poorly conditioned" if the number $\left| \frac{x_0 f'(x_0)}{f(x_0)} \right|$ is too large.

In particular, the relative error increases considerably when we calculate the values of a polynomial in the vicinity of its root. It is clear that in such a calculation we inevitably "eliminate significant digits".

Theorem 2. *The estimate of the absolute error of the algebraic sum of two functions does not exceed the sum of the estimates of the errors of the terms.*

Proof. Indeed, the differential of the sum of two functions is equal to the sum of the differentials of the terms and, consequently, the absolute value of the error of the sum does not exceed the sum of the absolute values of the absolute errors of the terms.

Note that Theorem 2 can be applied to the sum of any

functions whereas for definite functions the estimate it gives may prove to be too high. Thus, for instance, for the function $y = x^2 - 2x$ for $x = 1.324 \pm 0.0005$, the estimation of the errors of the terms yields

$$\delta_1 = 2 \cdot 1.324 \cdot 0.0005 \cong 0.001$$

for the first term $x^2 = 1.324^2 \pm \delta_1$ and

$$\delta_2 = 2 \cdot 0.0005 = 0.001$$

for the second term $2x = 2 \cdot 1.324 \pm \delta_2$. The estimation made according to Theorem 2 yields $| \Delta y | \leqslant 0.002$, and, if we make estimation according to Theorem 1, we have $\Delta y \cong (2x_0 - 2)\, \Delta x = 0.648 \Delta x$ and

$$| \Delta y | \leqslant 0.648 \cdot 0.0005 \cong 0.0003,$$

i.e. the estimate proves to be six times better.

Theorem 2 is true for the sum of two arbitrary functions and, therefore, it can be considered to be true for the sum of any approximate numbers. Namely, the following theorem is true.

Theorem 2'. *The estimate of the absolute error of the algebraic sum of two approximate numbers does not exceed the sum of the estimates of the errors of the terms.*

We can easily prove this theorem directly, without referring to Theorem 2.

Proof. Assume that u and v are two approximate numbers, u_0 and v_0 are their approximate values and Δu and Δv are their absolute errors. Then $u = u_0 + \Delta u$, $v = v_0 + \Delta v$ and $u + v = u_0 + v_0 + \Delta u + \Delta v$, whence the absolute error of the sum $u + v$ is $\Delta u + \Delta v$ and its absolute value does not exceed $| \Delta u | + | \Delta v |$, and this yields the required estimate.

However, in a certain sense, the estimate given by Theorem 2 is also too high. To understand this, consider an example. Let $u = 1.116 \pm 0.0005$, $v = 2.162 \pm 0.0005$. The estimate of the error of the sum is equal to ten units of the fourth decimal digit. Let us consider in more detail how the error of the sum appears. The notation $u = 1.116 \pm 0.0005$ means that the error of the quantity u may be equal to one of the eleven values

$$5 \cdot 10^{-4},\ 4 \cdot 10^{-4},\ 3 \cdot 10^{-4},\ 2 \cdot 10^{-4},\ 0,$$
$$-2 \cdot 10^{-4},\ -3 \cdot 10^{-4},\ -4 \cdot 10^{-4},\ -5 \cdot 10^{-4}$$

with an accuracy to within the fourth digit, and there is no reason to prefer one of these values to another. The same is true of the quantity v. Thus, for Δu and Δv there are $11^2 = 121$ possibilities with an accuracy to within the fourth digit, and in only two of them,

$$\Delta u = 5 \cdot 10^{-4}, \quad \Delta v = 5 \cdot 10^{-4}$$

and

$$\Delta u = -5 \cdot 10^{-4}, \quad \Delta v = -5 \cdot 10^{-4},$$

the error $10 \cdot 10^{-4}$ of the sum, maximum in absolute value, is realized. The following table gives the number of cases in which one or another absolute value of the error of the sum is realized.

Absolute value of the error of a sum ($\times 10^{-4}$)	10	9	8	7	6	5	4	3	2	1	0
Number of cases	2	4	6	8	10	12	14	16	18	20	11

It can be seen from the table that the absolute value of the error of $u + v$ does not exceed $5 \cdot 10^{-4}$ in 91 cases out of 121, does not exceed $7 \cdot 10^{-4}$ in 109 cases out of 121, and the maximum error equal to $10 \cdot 10^{-4}$ is realized in two cases and so is highly improbable. When we sum up a larger number of approximate values, the realization of the maximum, or close to maximum, error of the sum becomes even more improbable.

Theorem 3. *The estimate of the relative error of the product and the quotient of two functions is equal to the sum of the estimates of the relative errors of the factors or of the dividend and divisor respectively.*

Proof. Calculating, for the functions $y = uv$ and $z = u/v$, the absolute errors Δy, Δz as differentials, we obtain

$$\Delta y = u\,\Delta v + v\,\Delta u, \quad \Delta z = \frac{v \cdot \Delta u - u \cdot \Delta v}{v^2},$$

whence we have

$$\frac{\Delta y}{y} = \frac{u \cdot \Delta v + v \cdot \Delta u}{uv} = \frac{\Delta v}{v} + \frac{\Delta u}{u},$$

$$\frac{\Delta z}{z} = \frac{v \cdot \Delta u - u \cdot \Delta v}{v^2} \cdot \frac{v}{u} = \frac{\Delta u}{u} - \frac{\Delta v}{v}.$$

Passing to the estimates of the absolute values, we obtain the required values.

Thus the relative errors do not increase considerably when we perform the operations of multiplication and division.

The estimate in Theorem 3 is true for the arbitrary functions u and v, and, therefore, the following theorem is true.

Theorem 3′. *The estimate of the relative error of the product and the quotient of two approximate values is equal to the sum of the estimates of the relative errors of the factors and the dividend and divisor respectively.*

We can easily prove this theorem directly, without referring to Theorem 3.

It should be pointed out in conclusion that if we have a function of several arguments $w = F(u, v, \ldots)$ such that we can express dw in terms of du, dv, . . ., considering u, v, . . . to be functions of a certain parameter, then, knowing the approximate values u_0, v_0, . . . and the estimates of the absolute errors Δu, Δv, . . ., we can estimate the absolute error of the result of calculations replacing du, dv by Δu, Δv, . . . and passing to the estimate of the absolute values.

Indeed, formulas which yield the expressions for dw in terms of du, dv, . . . do not depend on what functions of the parameter the variables u, v, . . . are. We can consider these functions to be arbitrary and du, dv, . . . to be unconnected.

Example 1. $y = \dfrac{u+2v}{u+v+1}$, $u = 1.35 \pm 0.005$, $v = 2.21 \pm 0.005$. Find the approximate values of y and estimate the error.

Solution. We begin with the estimation of the error. We have

$$dy = \frac{(u+v+1)(du+2dv)-(u+2v)(du+dv)}{(u+v+1)^2} = \frac{(1-v)\,du+(u+2)dv}{(u+v+1)^2},$$

whence

$$\Delta y \cong \frac{-1.21\Delta u+3.35\Delta v}{(1.35+2.21+1)^2} \cong \frac{3.35\Delta v-1.21\Delta u}{21}$$

and

$$|\Delta y| \leqslant \frac{3.35}{21}|\Delta v| + \frac{1.21}{21}|\Delta u| \leqslant \frac{4.56}{21}\cdot 0.005 \cong 0.001.$$

This means that we should calculate the result with three decimal digits. It will deviate from the true result by not more than one unit of the third decimal digit.

Thus we have

$$y = \frac{1.35 + 4.42}{1.35 + 2.21 + 1} \pm 0.001 = 1.265 \pm 0.001.$$

EXERCISES TO CHAPTER 3

1. Prove the monotonicity of the given function $f(x)$ on the indicated intervals first by elementary means, showing that the inequality $x_2 > x_1$ implies the inequality $f(x_2) > f(x_1)$ or $f(x_2) < f(x_1)$, and then by the means of differential calculus.

1. $f(x) = \frac{3x-2}{x+1}$ decreases on $(-\infty, -1)$ and on $(-1, \infty)$.
2. $f(x) = \frac{3-2x}{x-2}$ increases on $(-\infty, -2)$ and on $(-2, \infty)$.
3. $f(x) = x^2 - 6x + 3$ increases on $[3, \infty)$.
4. $f(x) = x^2 + 8x - 7$ decreases on $(-\infty, -4]$.

2. Prove that if $g(x)$ increases on $(a, b]$ and on $[b, c)$, then $g(x)$ increases on (a, c).

3. Knowing the character of monotonicity of the functions $f(x)$ and $g(x)$, it is easy to find the character of monotonicity of the composite function composed of the functions $f(x)$ and $g(x)$, i.e. the functions

$$F(x) = f(g(x)) \text{ and } G(x) = g(f(x)),$$

using a law resembling the law of signs employed in multiplication, where a decreasing function plays the part of the minus sign and an increasing function plays the part of the plus sign. For instance, if $f(x)$ is a decreasing function and $g(x)$ is an increasing function, then $F(x)$ and $G(x)$ are decreasing functions; if $f(x)$ and $g(x)$ are decreasing functions, then $F(x)$ and $G(x)$ are increasing functions.

Prove these laws by elementary means and by the means of differential calculus on the assumption that $f(x)$, $g(x)$, $F(x)$ and $G(x)$ are differentiable functions.

4. Determine the character of monotonicity of each of the following functions:

1. $y = 2^{3x-4}$.
2. $y = \left(\frac{1}{2}\right)^{3-x}$
3. $y = (0.7)^{1-2x}$.
4. $y = (\sqrt{2})^{\frac{1}{3}x+1}$.
5. $y = \sqrt{6 - \frac{1}{3}x}$.
6. $y = \sqrt{2x - 0.5}$.
7. $y = 2 - \frac{3}{x}$.
8. $y = \frac{6}{x} + 5$.

5. Find the intervals of monotonicity of the following functions using the tests for the increase and decrease of functions:

1. $u(t)=t^3-t^2-t+7$. 2. $v(x)=\frac{1}{3}x^3-x^2-8x+3$.

3. $V(\alpha)=\frac{\alpha}{\alpha^2+4}$. 4. $P(x)=\frac{x^2}{x-3}$. 5. $f(x)=x\cdot\ln x$.

6. $g(x)=x\cdot e^x$. 7. $G(x)=\frac{\sqrt{x-3}}{x^2-4}$. 8. $F(t)=\frac{-\sqrt{5-t}}{t^2}$.

6. Find the intervals of monotonicity and the extrema of the following functions. Investigate the behaviour of these functions as $x\to\infty$ and in the vicinity of the points of discontinuity provided that they exist. Sketch the graphs of the functions:

1. $y=(x-1)(x-2)(x-3)$. 2. $y=x(x+2)(x-3)$.

3. $y=\frac{x^2+1}{1-x^2}$. 4. $y=\frac{x^2-2x+2}{2x-x^2}$. 5. $y=e^{x^2}$.

6. $y=(\ln x)^2$. 7. $y=x^2-2|x|$. 8. $y=-x^2+|x-2|$.

9. $y=\frac{x}{\sqrt{x^2+1}}$. 10. $y=\frac{x^2}{\sqrt{x^2+1}}$. ◀ 11. $y=x^x$, $x>0$.

12. $y=x^{1/x}$, $x>0$. ▶

7. Construct the graph of the function

$$f(x)=\frac{1}{5}x^5-\frac{1}{4}x^4-\frac{1}{3}x^3+\frac{1}{2}x^2+1$$

on the interval $|x|\leqslant 2$.

8. Construct the graphs of the following curves:

1. $f(x)=e^{-x^2}$. 2. $f(x)=e^x-x+1$. 3. $y=\sqrt{9x^2-1}$.

4. $y=\sqrt{16-x^2}$. 5. $y=0.2x^3+0.3x^2-1.2x+0.1$.

6. $y=0.1x^4-0.4x^3+0.4x^2+0.5$. 7. $y=\frac{x^2-x}{x^2+1}$.

8. $y=\frac{x^2-x+1}{x^2-x-2}$. 9. $y=\ln\frac{1+x}{2-x}$. 10. $y=\ln\frac{x+1}{x-2}$.

11. $y=\sqrt[3]{x^2-16}$. 12. $y=\sqrt[3]{8-x^3}$. 13. $y=x^{\sqrt{x}}$.

14. $y=x^{m/n}$, where m and n are coprime natural numbers, $x\geqslant 0$.

9. Using the methods of differential calculus, investigate the behaviour of the function $y=ax^2+bx+c$, where a, b and c are constants.

10. Find the greatest value of the function $f(x)$:

1. $f(x)=x^2(1-x)^3$, $0\leqslant x\leqslant 1$.
2. $f(x)=x^p(a-x)^q$, $0\leqslant x\leqslant a$; p and q are positive constants.

11. Find the greatest and the least value of each of the following functions on the indicated intervals:

1. $f(x) = x^3 - 12x + 4, \ -3 \leqslant x \leqslant 4.$
2. $g(x) = t^4 - 4t^3 - 8t^2 + 13, \ -2 \leqslant t \leqslant 3.$
3. $V(\alpha) = \alpha^4 - 9\alpha^2 + 4, \ 0 < \alpha < 6.$
4. $P(h) = h^3 - 5h^2 + 7h - 3, \ 0 < h < 2.$
5. $f(x) = \ln(x^2 - 2x - 1), \ -3 \leqslant x \leqslant -1.$
6. $g(x) = \sqrt{x^2 - 4x + 2}, \ 4 \leqslant x \leqslant 6.$

12. Prove that if the sum of two positive numbers is defined, then their product assumes the greatest value when these numbers are equal.

13. Assume that we have some water whose volume at the temperature $t = 0°$ is unity. According to the experimental data, the volume $v(t)$ of this quanitity of water changes with the rise of temperature t according to the law

$$v(t) = 1 - 0.61045 \cdot 10^{-4}t + 0.77183 \cdot 10^{-5}t^2 - 0.373 \cdot 10^{-7}t^3.$$

Find the temperature at which the volume of water is the least. What can you say about the density of water at this temperature?

14. According to Ohm's law for a closed circuit, the electric current is proportional to the electromotive force E of the source and inversely proportional to the sum of the resistance of the external circuit R and the resistance of the source r, i.e.

$$I = \frac{E}{R + r}.$$

Consequently, varying the resistance of the external circuit R, we can vary the current I. Show that the power supplied by the source is the greatest when the resistance of the external circuit R is equal to the internal resistance of the source r.

15. Prove that of all isosceles triangles inscribed in a given circle, a regular triangle has the greatest area.

16. A rectangle is inscribed in a triangle with base a and altitude h. One side of the rectangle lies on the base of the triangle and two vertices lie on its lateral sides. Find the greatest area of the inscribed rectangle.

17. Two points move along the coordinate axes according to the laws $x(t) = 2t - 9$ and $y(t) = -3t + 7$ $(t \geqslant 0)$. At what moment will the distance between the points be the least? Depict the position of the points at the moment $t = 0$ and at the moment of their maximum approach to each other.

18. Let $F(x) = f(g(x))$, where the function $f(x)$ is increasing (decreasing) and the function $g(x)$ assumes the greatest value at the point x_0 belonging to the domain of definition of the function $F(x)$. Show that $F(x)$ assumes the greatest (the least) value at the same point x_0. What can you say about the value of the function $F(x)$ at the point x_0 if the function $g(x)$ has the least value at this point?

19. Use the results of the preceding exercise to find the greatest and the least value of the following functions on the indicated intervals:

1. $f(x) = \ln(-x^2 + 10x + 1)$, $0 \leqslant x \leqslant 6$.
2. $g(x) = \left(\frac{1}{2}\right)^{x^2+4}$, $-2 \leqslant x \leqslant 1$.
3. $S(t) = \sqrt{t^2 + 4t + 7}$, $-3 \leqslant t \leqslant 0$.
4. $V(h) = 3 \cdot (h^2 - 2h + 2)^{-1}$, $0 \leqslant h \leqslant 6$.
5. $P(\alpha) = 2^{5-\alpha^2}$, $-2 \leqslant \alpha \leqslant 1$.
6. $H(l) = \log_{0.7}(l^2 + 6l + 11)$, $-4 \leqslant l \leqslant -1$.

20. A particle moves in a straight line according to the law $x(t) = \frac{1}{2}t^3 - t^2 + 3t - 6$, where x is the coordinate of the particle on the straight line and t is time. Show that the acceleration of the particle varies according to a linear law.

21. Use successive differentiation to derive Leibniz's formulas:

1. $(u \cdot v)'' = u''v + 2u'v' + uv''$.
2. $(u \cdot v)''' = u''' \cdot v + 3u''v' + 3u'v'' + uv''$.
3. $(u \cdot v)^{(4)} = u^{(4)}v + 4u'''v' + 6u''v'' + 4u'v''' + uv^{(4)}$.
4. $(u \cdot v)^{(5)} = u^{(5)} \cdot v + 5u^{(4)} \cdot v + 10u''' \cdot v'' + 10u'' \cdot v''' + 5u' \cdot v^{(4)} + u \cdot v^{(5)}$.

22. Use Leibniz's formulas to find the derivatives of the second and the third order of the following functions:

1. $f(x) = e^x x^3$. 2. $f(x) = e^x/x$.
3. $g(x) = x^3 \ln(-x)$. 4. $g(x) = x^2 \ln x$.
5. $f(x) = (x - 3)^4 (x + 5)^3$. 6. $f(x) = \ln x^{x^2+4}$.

23. For each of the given functions find the highest order of the derivative which is not identically zero:

1. $y = 3x^4 - 7x^3 + 6$. 2. $y = (1 - x - x^2)^3$.
2. $y = (2x - 1)^2 (1 - 3x)^3$.
4. $y = (x - 1)(x - 2)(x - 3)(x - 4)$.

24. Find the derivatives of the function $y = \frac{1}{2}\left(\frac{1}{1-x} + \frac{1}{1+x}\right)$ up to the fifth order inclusive. Calculate the values of the derivatives for $x = 0$.

25. Obtain formulas for abridged multiplication for the square and the cube of the sum of two numbers using the binomial theorem.

26. Calculate the binomial coefficients for the fifth power of the binomial, i.e. calculate $\binom{5}{k}$ for all k from 0 to 5. Write the formula for raising the sum of two terms to the fifth power.

27. Prove the identity $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$.

28. Follow the way the identity from the preceding problem is used when the table of binomial coefficients, i.e. *Pascal's triangle*, is compiled. Working independently, extend the table to exponent 10.

n \ k	0	1	2	3	4	5	6	7	8	9	10
0	1										
1	1	1									
2	1	2	1								
3	1	3	3	1							
4	1	4	6	4	1						
5	1	5	10	10	5	1					
6											
7											
8											

29. Use Pascal's triangle to raise the following expressions to the indicated power:

1. $(1+x)^8$. 2. $(1+x)^{10}$. 3. $(1-x)^9$.
4. $(1-x)^7$. 5. $\left(\frac{1}{2}+2x\right)^6$. 6. $\left(\frac{1}{2}x-2\right)^5$.
7. $(\sqrt{x}-\sqrt{y})^4$. 8. $(a+b)^{10}$. 9. $(a-b)^{10}$.

30. Using the binomial theorem for $(1+x)^n$, show that:

1. $\binom{n}{0}+\binom{n}{1}+\binom{n}{2}+\ldots+\binom{n}{k}$
$$+\ldots+\binom{n}{n-1}+\binom{n}{n}=2^n.$$

2. The sum of the binomial coefficients that occupy the even places is equal to the sum of the binomial coefficients that occupy the odd places.

Verify the validity of these assertions by way of particular examples using Pascal's triangle.

31. Compare the numbers $(1.02)^5$ and $(1.03)^4$.

32. Establish the relationship between the shape of the graph of a function (convexity upwards, convexity downwards) and the character of the monotonicity of the derivative of the function by constructing tangents to the graph of the function at several points and fol-

lowing the variation of the angle of inclination of a tangent with an increase in the argument x (Figs. 3.24 and 3.25).

33. Draw the graph of the function $f(x) = x^3$. Explain the properties of the derivative of this function from the graph. Verify your

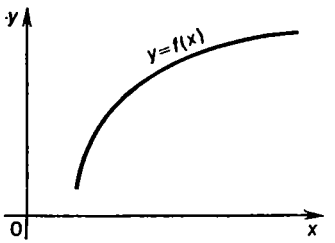


Fig. 3.24

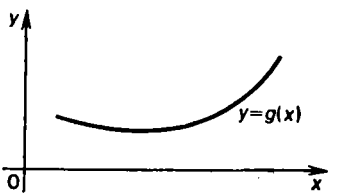


Fig. 3.25

inferences by carrying out the exact investigations of the derivative of the function $f(x) = x^3$.

34. Figure 3.26 shows parts of the graphs of the twice differentiable functions $f(x)$ and $g(x)$. Here x_0 is the abscissa of the lowest point of the graph of $f(x)$ and x_1 is the abscissa of the uppermost point of the graph of $g(x)$.

Make similar drawings. Having performed the necessary constructions, follow the variation of the angle of inclination of the tangent

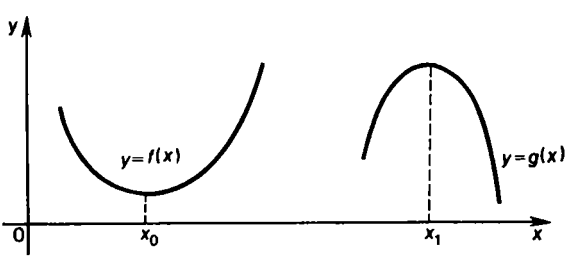


Fig. 3.26

to the graph of the function $f(x)$ in the vicinity of the point x_0. Make sure that $f'(x)$ increases in the vicinity of the point x_0 changing sign from minus to plus at the point x_0, being given that $f'(x_0) = 0$ and $f''(x_0) > 0$. In the same way make sure that the function $g'(x)$ decreases in the vicinity of the point x_1 changing sign from plus to minus at the point x_1, being given that $g'(x_1) = 0$ and $g''(x_1) < 0$.

35. First, using the graph of the function $y(x) = x^4 - 2x^2$ (3.3, Fig. 3.9), try to find the intervals of monotonicity of the derivative of $y(x)$ and determine the inflection points by eye. Then verify the results obtained by finding the exact solution of the problem.

36. Figure 3.27 shows the graph of the movement of a particle in a straight line, $x(t)$ being the coordinate of the point on the straight line at the moment t. Which of the parts of the graph corresponds to the motion with positive acceleration and which to the motion with negative acceleration? Mark the moment at which the acceleration is zero.

37. Construct the graph of the function $y = y(x)$ and find the inflection points and the direction of the convexity:

1. $y = xe^x$. 2. $y = (x+4)\sqrt[3]{x}$. 3. $y = x^3 - 4x^2 - 3x + 6$.

4. $y = -x^3 + 6x^2 - 9x$. 5. $y = (x^2-1)^3$. 6. $y = \dfrac{(x-1)^2}{(x+1)^3}$.

38. Give substantiated answers to the following questions, and, if possible, illustrate the answers by graphs:

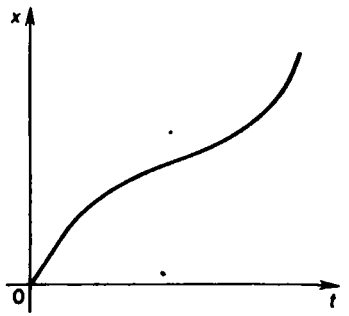


Fig. 3.27

1. Can a function defined on an interval, whose graph is convex upwards or convex downwards throughout the interval,
   (a) have no extrema?
   (b) have only one extremum (what extremum?)?
   (c) have more than one extremum?
   (d) cut a straight line more than twice?
   (e) have more than one point in common with any one of its tangents?
2. Can a function which has no extrema change the direction of convexity?
3. What is the position of a curve relative to its tangent in the vicinity of the point of tangency if the latter is an inflection point at the same time?

39. Using the graph of the function shown in Fig. 3.28, where a_1, a_2, a_3 are points of extremum and b_1, b_2, b_3 are points of inflection, sketch the graphs of its first and second derivatives.

40. Prove that the proximity of the function $y = \frac{1}{4}x^4 - \frac{5}{6}x^3 + x^2 - x$ to its tangent at the point with abscissa $x_0 = 1$ as $x \to 1$ is 3

41. Prove that the proximity of the function $y = \sqrt{x}$ to its tangent at the point with abscissa $x_0 = 4$ as $x \to 4$ is of order 2.

42. For the function $y = x^3$ choose a quadratic trinomial which would have a proximity of order 3 in the vicinity of the point $x = 1/2$.

43. Prove that the order of smallness of the function $f(x) = e^x - \frac{1}{2}x^2 - x - 1$ in the neighbourhood of the point $x = 0$ is 3.

44. Prove that if a polynomial is divisible by the expression $(x-\alpha)^n$, then the order of smallness of the polynomial in the vicinity of $x = \alpha$ is equal to or higher than n.

45. Prove a statement converse to that formulated in the preceding problem.

46. Proceeding from the statement from Problem 44 and the converse statement, find the divisor of the form $(x - \alpha)^n$ for the polynomial

$$P(x) = 0.1x^5 - 0.25x^4 - 2x^3 - 3.5x^2 - 2.5x - 0.65.$$

Factor the polynomial $P(x)$.

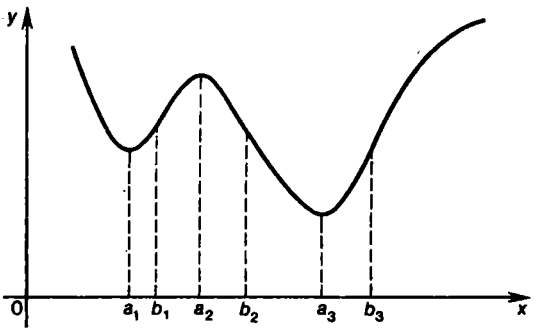


Fig. 3.28

47. Find the polynomial $P(x)$ of the third degree that would have a proximity of order 4 to the function $y = \ln x$ in the neighbourhood of the point $x = 1$.

48. Expand the polynomial

$$P(x) = x^4 - 3x^3 + 6x^2 - 10x + 2$$

in the powers of the binomial $(x - 2)$. Calculate $P(2.02)$ with an accuracy of 10^{-4}.

49. Write the equation of the tangent to the graph of the polynomial

$$P(x) = x^4 - 3x^3 + 6x^2 - 10x + 2$$

at the point with abscissa $x_0 = 2$.

How can this equation be obtained from the expansion of $P(x)$ in the powers of $(x - 2)$? What is the proximity of the tangent to the graph of $P(x)$ in the vicinity of $x = 2$?

50. Use Maclaurin's formula to write the expansion of the function $f(x) = \sqrt{1 + x}$ to the term containing x^3. Calculate $\sqrt{1.02}$ with an accuracy of 10^{-4}.

51. A rounded value of the number π with seven decimal digits is 3.1415926. Write a rounded value of the number π with two, three, four and five decimal digits. In each case estimate the absolute round-off error.

52. The mass of a component of a drug (in grammes), measured by means of a medical balance is $m = 15 \pm 0.2$. The mass of a car (in kg), measured by means of a truck scales, is $M = 43200 \pm 10$. Com-

pare the accuracy of measurements using the estimate of the relative error.

53. Employing the theorem on the constancy of a relative error in the multiplication by an exact number, find the relative error of the approximate values of the following quantities:

$$a = 458,\ b = 45.8,\ c = 4.58,\ d = 0.458,$$

assuming that all the digits in the notations of the approximate values are correct and the next digits are unknown. Express the error in percent.

54. We have five measurements of the same quantity with an increasing number of valid significant digits:

$$3,\ 3.1,\ 3.14,\ 3.141,\ 3.1415.$$

Follow the variation of the relative error with an increase in the number of valid significant digits.

55. Let $x = 3.14 \pm 0.002$. Using a calculator, we have found that $\sqrt{3.14} = 1.7720045$. Estimate the absolute error of the value of the square root obtained. How many valid decimal digits are there in the answer?

56. Solve Problem 55 on the assumption that $x = 2.7 \pm 0.02$ and $\sqrt{2.7} = 1.6431676$.

57. Use a calculator to find the value of a function for a given approximate value of the argument in the cases given below. Estimate the absolute and the relative error using Theorem 1 from 3.12.

1. $f(x) = \ln x, \quad x = 5.2 \pm 0.05.$
2. $f(x) = e^x, \quad x = 2 \pm 0.2.$
3. $f(x) = \dfrac{x}{x+7}, \quad x = -3.4 \pm 0.05.$
4. $f(x) = 3x^2 + 4x, \quad x = -7.32 \pm 0.005.$

58. Write the given approximate values of the quantities in the form which uses the mantissa and the exponent of the number, retaining only valid digits:

1. $x = 27.5 \pm 0.05.$
2. $x = 48 \pm 0.5.$
3. $x = 0.0412 \pm 0.00002.$
4. $x = 0.57 \pm 0.004.$
5. $x = 401200 \pm 50.$
6. $x = 12084000 \pm 100.$
7. $x = 42.0 \pm 0.2.$
8. $x = 27.0 \pm 0.01.$

59. Find the absolute and the relative error of the approximate values whose notations include only valid digits:

1. $x = 3.4.$
2. $x = 0.872.$
3. $x = 5.65 \cdot 10^{-3}.$
4. $x = 2.5 \cdot 10^{-5}.$
5. $x = 7.546 \cdot 10^{6}.$
6. $x = 8.025 \cdot 10^{8}.$
7. $x = 2.60 \cdot 10^{-2}.$
8. $x = 4.0 \cdot 10^{3}.$

60. The measurement of the length of two segments yielded the following results:

$$l_1 = 14.3 \pm 0.05 \text{ (cm)}, \quad l_2 = 14 \pm 0.5 \text{(cm)}.$$

Find the approximate value of the sum of the lengths of these segments. Can we determine which of the segments is longer?

61. Find the sum of the approximate values of the given numbers. Estimate the absolute error:

1. $S = x + y + z$, where $x = 34.1 \pm 0.02$, $y = 7.13 \pm 0.005$, $z = 11.5 \pm 0.05$.

2. $M = a + b - c - d$, where $a = 3.427 \pm 0.0003$, $b = 4.845 \pm 0.0005$, $c = 1.453 \pm 0.0005$, $d = 2.602 \pm 0.0002$.

62. Find the value of each of the given expressions and estimate the relative errors of the results:

1. $P = \dfrac{xy}{z}$, where $x = 14.2 \pm 0.05$, $y = 0.25 \pm 0.005$, $z = 0.000862 \pm 0.0000005$.

2. $Q = \dfrac{a}{b \cdot c}$, where $a = 2.73 \cdot 10^5 \pm 5 \cdot 10^2$, $b = 4.7 \cdot 10^2 \pm 5$, $c = 6.07 \cdot 10^4 \pm 5 \cdot 10^1$.

63. Use a calculator to find the value of each of the following expressions for the given approximate values of the letters entering into it. Estimate the absolute error of the calculations regarding the initial expression as a function of several variables:

1. $y = \dfrac{\alpha^2 - 3\beta}{2\alpha + \beta}$, $\alpha = 4.15 \pm 0.01$, $\beta = 1.25 \pm 0.01$.

2. $y = \dfrac{u - v - v^2}{u^2 + 4v}$, $u = 4.2 \pm 0.1$, $v = 6.7 \pm 0.1$.

Additional Exercises

64. At what points of the parabola $y = x^2$ do the tangents to it cut the x-axis at an angle of 150°, 135°, 120°, 60°, 30°?

65. At what points of the cubic parabola $y = \dfrac{1}{8} x^3$ do the tangents to it cut the abscissa axis at an angle of 135°, 45°?

66. At what points of the curve $y = \dfrac{x^3}{3} - x^2 + \dfrac{1}{3}$ are the tangents parallel to the x-axis? At what points of the curve do the tangents cut the x-axis at an angle of 135°?

67. Find the angles of inclination of the tangents to the curve

$$f(x) = x^3 - 6x^2 + 11x - 6$$

at the points where $f(x) = 0$.

68. Show that the tangents to the curve $f(x) = \dfrac{x}{1 - x^2}$ at the points

$$(0, 0), \left(\sqrt{3}, -\frac{1}{2}\sqrt{3}\right), \left(-\sqrt{3}, \frac{1}{2}\sqrt{3}\right)$$

are parallel and cut the x-axis at an angle of 45°. Draw the graph of the function $f(x)$.

69. Find the slope of the tangent to the curve

$$x^3 - 7x^2y - 5y^3 + 4x^2 - 10xy + 8x - 5y + 18 = 0$$

at the point (2.1).

70. Write the equation of the tangent to the graph of the function

$$y = \frac{1}{2}x^2 - 3x + 7$$

which makes an angle of 45° with the abscissa axis.

71. Find the points of intersection of the curves

$$x^3 + y^2 - 8 = 0 \quad \text{and} \quad y^2 - 2x = 0.$$

Construct tangents to the curves at the points of their intersection. Use a protractor to determine the *angles of intersection* of the curves as the angles between the tangents.

72. Prove that $x^a \ln x$, $a > 0$, is bounded on the interval $(0, 1]$. Prove that $x^a \ln x \to 0$ as $x \to 0$.

73. Construct the graphs of the parabolas $y^2 - 4x = 0$ and $x^2 - 4y = 0$. Write the equations of the tangents to these curves at the points of their intersection and sketch them. Using appropriate tables, find the angles of intersection of the curves (tangents).

74. Construct the graphs of the hyperbolas $x^2 - y^2 = 8$ and $xy = 4$. Write the equations of the tangents to these curves at the points of their intersection and sketch them. Use appropriate tables to find the angles of intersection of the curves (tangents).

75. Show that the function $f(x)$ has only one root and find it?

1. $f(x) = (x-2)e^x + x + 2$.

2. $f(x) = \frac{x^2}{2e} - \ln x$. 3. $f(x) = e^x - \sqrt{2e}\sqrt{x}$.

76. Prove that the function $f(x)$ is monotonic on each of the following intervals:

1. $f(x) = 2\ln\frac{x}{a+x} + \frac{a}{x} + \frac{a}{a+x}$, $x \in (0, \infty)$ $a > 0$.

2. $f(x) = \left(\frac{x}{a+x}\right)^{a+2x}$, $x \in [0, \infty)$, $a > 0$.

77. Construct the graph of the function $f(x) = e^x - 5x + 1$. Solve the equation $f(x) = 0$ by graphical means.

78. It has been proved by experiments that when a body falls freely in water the resistance of the water grows in proportion to the rate of the immersion and the path traversed by the body can be found from the formula

$$S(t) = \frac{g}{k}t + \frac{g}{k^2}(e^{-kt} - 1), \quad t \geqslant 0,$$

where k is a positive constant and g is the acceleration of free fall. Show that the rate $v(t)$ of the immersion of the body and its accelera-

tion $a(t)$ satisfy the relation $a(t) = g - kv(t)$. Is it true that beginning with a certain moment the body falls with almost constant velocity.

79. A particle moves in a straight line so that at any moment t the quantities $x(t)$ and $v(t)$ are related as $v^2(t) = 2gx(t)$, where g is a constant. Show that the particle moves with constant acceleration.

80. A point moves in a straight line and the coordinate of its position $x(t)$ at the moment t is defined by the formula $x(t) = 4t^3 - 21t^2 + 36t + 1$, t is measured in seconds ($t \geqslant 0$). Find the speed $v(t)$ and the acceleration $a(t)$ of the moving point. Find out when the point moves to the left of the abscissa axis and when to the right. Find the moments when the point stops. Determine the intervals of time when the point moves with positive acceleration and with negative acceleration.

81. Mark points on the coordinate plane whose coordinates are given in the table. The scale: a unity is equal to a side of a square in your copybook.

x	0	1	2	4	6	9	10	11	12	13	14	15	16	17	19	22
y	0	2	3.5	6	8	10	9	7	5.5	5	6	8	10	11	12	13

Connect the constructed points by a continuous line. Let this line be the graph of the function $y = f(x)$. In the drawing mark the points at which the derivative $f'(x)$ vanishes. Mark the intervals where $f'(x) > 0$ and $f'(x) < 0$. Noting how the slope of the tangent to the graph of the function changes with the movement of the point of tangency along the graph, sketch the graph of the derivative $f'(x)$ in the same coordinate system.

82. Construct curves symmetric to the graph of the function $y = f(x)$ from the preceding exercise relative to:

(1) the axis of ordinates, (2) the abscissa axis, (3) the origin.

Carry out the assignments from Problem 81 for each of the curves obtained.

83. We have the coordinates of the vertices of the polygonal line $ABCD$, which is the graph of the derivative $f'(x)$ of a function $f(x)$, and the value of the function $f(x)$ at one of the points. Sketch the graph of the initial function $f(x)$.

1. $A(-8, -4)$, $B(-2, 6)$, $C(6, -2)$, $D(12, -2)$, $f(-8) = 10$.
2. $A(-10, 3)$, $B(-3, 3)$, $C(4, 4)$, $D(8, 1)$, $f(-10) = -6$.

84. Choose several functions which would possess the given derivative:

1. $f'(x) = 2x$. 2. $f'(x) = 5$. 3. $f'(x) = 1/x$.

4. $f'(x) = \dfrac{1}{2\sqrt{x}}$. 5. $f'(x) = -\dfrac{1}{x^2}$. 6. $f'(x) = e^x$.

Chapter 4

TRIGONOMETRIC FUNCTIONS

4.1. Generalization of the Concept of an Angle

Trigonometric functions play a significant part in mathematics and its applications. In this connection, it has become necessary to develop a powerful technique of transformation of expressions including trigonometric functions and the technique of solving corresponding equations. This technique is based on algebra, has much in common with algebra and is known as *trigonometry*. The use of elements of trigonometry in solving geometry problems is considered in geometry. Trigonometry has found still more important applications in analysis. The matter is that with the appropriate generalization of the concept of an angle trigonometric functions prove to be periodic, i.e. their values are repeated in certain intervals. In a sense they turn to be the simplest elementary functions. We come across periodic functions in descriptions of oscillatory motions such, for instance, as oscillations of a loaded spring, oscillations of a pendulum, oscillations of voltage in an a.c. circuit. Furthermore, as we know from physics, such phenomena as sound, light, radio waves are connected with oscillation processes and the apparatus of trigonometric functions is necessary for their study.

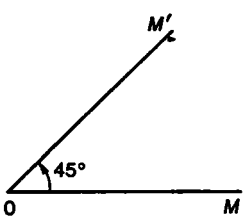


Fig. 4.1

First of all we must verify whether trigonometric functions are indeed periodic. To do this, we must generalize the concept of an angle.

Assume that a half-line OM originating at O can oscillate about its origin. Then, it is natural to measure the amount of rotation by the measure of the angle which is formed between the initial and the terminal position of the half-line. Thus, in Fig. 4.1 the half-line OM' is obtained as a result of the counterclockwise rotation of the half-line OM through an angle of 45°.

However, to be able to measure the amount of rotation by the measure of an angle, we must extend the concept of an angle assuming angles exceeding 180° to be admissible and meaningful. Thus it is natural to consider a rotation through

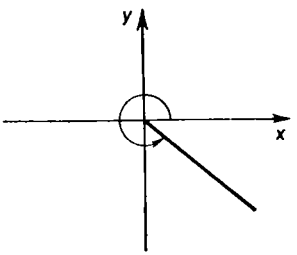


Fig. 4.2

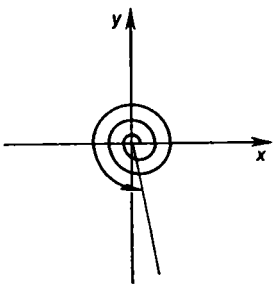


Fig. 4.3

3/4 of a complete revolution to be a rotation through 270°, a complete revolution to be a rotation through 360°, one and a half revolution to be a rotation through 540° and so on. It stands to reason that if, after a rotation through an angle φ_1 we make a rotation through an angle φ_2, then the angle

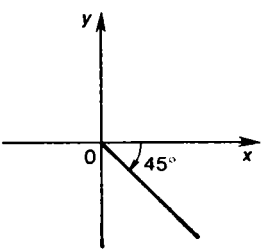


Fig. 4.4

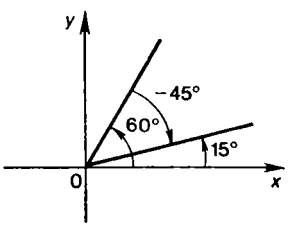


Fig. 4.5

measuring the rotation is equal to $\varphi_1 + \varphi_2$. Figures 4.2 and 4.3 show rotations through 315° and through 1000° respectively.

It is customary to consider counterclockwise rotation to be positive and clockwise rotation to be negative and the respective angle of rotation to be also negative. Figure 4.4 shows an angle of rotation through —45° and Fig. 4.5 shows

a rotation first through 60° and then through —45°. The resulting rotation is through 15°.

Of course, the choice of the positive direction of angle reckoning is conventional. For instance, it would be inconvenient to consider the hands of a clock to be moving in the negative direction if we assumed the clockwise direction to be positive. If we choose a coordinate system in the plane,

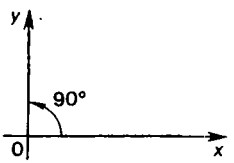


Fig. 4.6

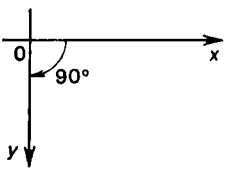


Fig. 4.7

then the positive direction of angle reckoning is the direction from the positive semi-axis of abscissas to the positive semi-axis of ordinates. For the coordinate axes in their usual position (Fig. 4.6), the counterclockwise direction is the positive direction of angle reckoning and for the position of the coordinate axes shown in Fig. 4.7 the clockwise rotation is positive.

4.2. Measuring Angles in Radians

Degree measure of angles is inconvenient for the purposes of analysis and so it is natural and convenient to measure angles in radians. The *radian measure* of a central angle is the ratio of the length of the arc subtending it to the radius of the circle. It follows from the theory of similitude that all figures bounded by a given angle and the arc of a circle with centre at the vertex of the angle are similar and so the radian measure of an angle does not depend on the radius of the circle.

It is clear that a right angle with the vertex at the centre of a circle cuts out of it a quadrant with the arc length of $\frac{1}{4} \cdot 2\pi r = \frac{\pi}{2} r$. Consequently, the radian measure of a right angle is $\frac{\pi}{2} r \div r = \frac{\pi}{2}$. The comparison of a radian and a degree measure of a right angle yields $\pi/2 = 90°$, whence it follows

that an angle of 1° has a radian measure $\pi/180°$. The radian measure of an angle of $\frac{180°}{\pi} = 57.29578° = 57°17'46''$ is equal to 1. This angle is known as a *radian*. A central angle equal to 1, is subtended by an arc whose length is equal to the radius of the circle. The radian measure of an angle is considered to be an abstract number, and this is natural since it is the ratio of lengths of two curves. We say that a right angle is equal to $\frac{\pi}{2}$ (and not to $\frac{\pi}{2}$ of a radian). Correspondingly, one radian is simply written as the number 1.

The radian measure is naturally extended to all angles larger than 180° and to negative angles. Thus one complete revolution in the positive direction is a rotation through an angle of 2π. The angle of $-900°$ has a radian measure -5π and so on.

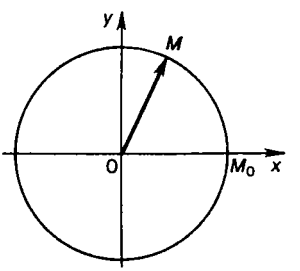


Fig. 4.8

Let us now interpret in more detail the radian measure for arbitrary angles considered to be angles of rotation. We choose a coordinate system in a plane and consider on it a circle with centre at the origin and a radius equal to 1. We assume the positive direction of the abscissa axis to be the initial position of the half-line which can rotate (Fig. 4.8). The position of the half-line extending from the origin is well defined by the position of the point M of intersection of the half-line and the circle. When the half-line rotates, the point M moves along the circumference of the circle. Assume that the half-line has rotated through an angle φ in the positive direction. Then, beginning with the position M_0, the point M will traverse a path along the circle. From the definition of the radian measure of an angle (understood in the generalized sense) and from the fact that a radius is equal to 1, it is clear that this measure is equal to the length of the path traversed by the point M. Thus, for $\varphi = 5\frac{1}{2}\pi$ the point M will make five and a half semirevolutions, i.e. will indeed traverse a path $5\frac{1}{2}\pi$ long. Now when the rotation is in the negative direction, the length of the

path traversed by the point M is equal to the modulus of the measure of the angle.

In what follows, we shall mainly use the radian measure of angles.

4.3. The Sine and Cosine Functions

Recall that the *sine* of the acute angle φ is the ratio of the length of the leg lying opposite to the angle φ to the length of the hypotenuse of a right triangle with acute angle φ. The ratio of the length of the leg adjacent to the angle φ to the length of the hypotenuse is the *cosine* of the angle φ. The sine and cosine of the angle φ are designated as $\sin \varphi$ and $\cos \varphi$ respectively. It follows from the theory of similitude that the values of $\sin \varphi$ and $\cos \varphi$ do not depend on the choice of a right triangle with acute angle φ since all triangles of this kind are similar.

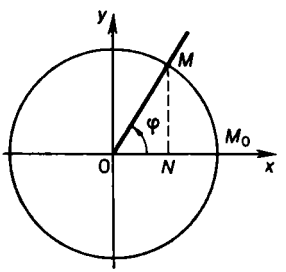


Fig. 4.9

Let us take a circle with centre at the origin and radius 1. We lay off an acute angle φ in the positive direction from the positive direction of the abscissa axis. We designate as M_0 and M, respectively, the points at which the positive direction of the abscissa axis and the second side of the angle we have constructed cut the circle (Fig. 4.9). We draw a perpendicular MN from the point M to the abscissa axis, and then the triangle OMN is right-angled with an acute angle φ at the vertex O and hypotenuse OM whose length is 1. Therefore the cosine and sine of the angle φ are equal to the lengths of the segments ON and MN respectively, i.e. to the abscissa and ordinate of the point M. This property of the cosine and sine established for acute angles can naturally be accepted as the definition of the values of these functions for any values of the angles in the generalized sense. By definition, the cosine and sine of the angle φ are, respectively, the abscissa and the ordinate of the point M at which the axis of abscissas, rotating through the angle φ in the positive direction, cuts the circle.

Thus, for instance, $\cos\frac{\pi}{2}=0$, $\sin\frac{\pi}{2}=1$, $\cos\pi=-1$, $\sin\pi=0$, $\sin 0=0$, $\cos 0=1$, $\cos\frac{7}{2}\pi=0$, $\sin\frac{7}{2}\pi=-1$.

The value of the argument of a function is always assumed to be an abstract number. It is quite natural to treat trigonometric functions as functions of a numerical argument. Thus the sine (cosine) of the number x is the sine (cosine) of an angle in x radians.

Let us calculate the values of the sine and cosine of some angles (numbers).

We consider an equilateral triangle ABC (Fig. 4.10). All of its angles are equal to $60^\circ = \pi/3$. We construct a bisector

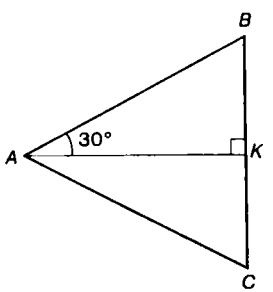


Fig. 4.10

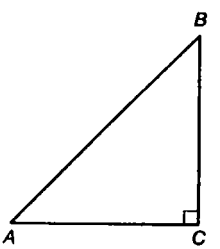


Fig. 4.11

AK of the angle A. It is a median and an altitude, and so $BK=\frac{1}{2}BC=\frac{1}{2}AB$ and $\angle AKB$ is a right angle. Consequently, $\sin 30^\circ=\frac{BK}{AB}=\frac{1}{2}$. Furthermore,

$$\cos 30^\circ=\frac{AK}{AB}=\frac{\sqrt{AB^2-BK^2}}{AB}=\sqrt{1-\frac{1}{4}}=\frac{\sqrt{3}}{2},$$

$$\sin 60^\circ=\frac{AK}{AB}=\frac{\sqrt{3}}{2};\quad \cos 60^\circ=\frac{BK}{BA}=\frac{1}{2}.$$

Let us consider now an isosceles right-angled triangle ABC with the right angle at the vertex C (Fig. 4.11). The angles at the vertices A and B are equal and are $45^\circ=\pi/4$ each. According to the Pythagorean theorem, the length of the

hypotenuse AB is $\sqrt{AC^2 + BC^2} = AC\sqrt{2}$. Consequently,

$$\sin 45° = \cos 45° = \frac{BC}{AB} = \frac{AC}{AB} = \frac{AC}{AC \cdot \sqrt{2}} = \frac{1}{\sqrt{2}} = \frac{\sqrt{2}}{2}.$$

◀ Let us calculate the values of the sine and cosine of the angles $18° = \frac{\pi}{10}$ and $72° = \frac{2}{5}\pi$. For this purpose, we consider an isosceles triangle ABC with the angle of 36° at the vertex A which lies opposite to the base of the triangle (Fig. 4.12). Let the lateral sides AB and AC be equal to 1.

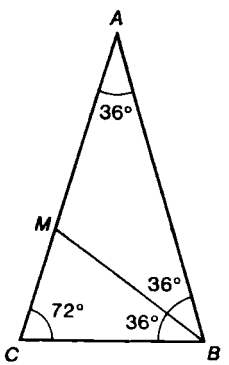


Fig. 4.12

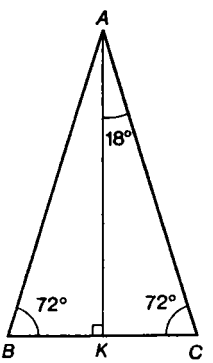


Fig. 4.13

The vertex angles B and C adjacent to the base are of 72°. We draw a bisector of the angle B. It divides the triangle ABC into two triangles MAB and BCM. The first triangle is isosceles since the vertex angles A and B are equal and are 36° each. The vertex angle M is $180° - 2 \times 36° = 108°$. The vertex angle M in the triangle BCM is $180° - 108° = 72°$. The vertex angle C in this triangle is of 72° and therefore the triangle BCM is isosceles. Thus $BC = BM = AM$ and $MC = AC - AM$. We designate the length of BC as x. Then $MC = 1 - x$. The triangles ABC and BCM are similar since the respective angles in them are equal. Equating the ratios of the respective sides, we obtain

$$\frac{x}{1} = \frac{1-x}{x},$$

whence it follows that $x^2+x-1=0$, $x=\frac{-1\pm\sqrt{5}}{2}$. The minus sign does not seem proper here and therefore $BC=x=\frac{1}{2}(\sqrt{5}-1)$. We draw an altitude AK (Fig. 4.13) and calculate its length

$$AK=\sqrt{AB^2-\left(\frac{BC}{2}\right)^2}=\sqrt{1-\left(\frac{\sqrt{5}-1}{4}\right)^2}=\frac{\sqrt{10+2\sqrt{5}}}{4}.$$

$$\sin 18^\circ=\frac{BK}{AB}=\frac{\sqrt{5}-1}{4},$$

$$\cos 18^\circ=\frac{AK}{AB}=\frac{\sqrt{10+2\sqrt{5}}}{4},$$

$$\sin 72^\circ=\frac{AK}{AB}=\frac{\sqrt{10+2\sqrt{5}}}{4},$$

$$\cos 72^\circ=\frac{BK}{AB}=\frac{\sqrt{5}-1}{4}. \blacktriangleright$$

4.4. The Fundamental Properties of the Sine and Cosine Functions

Theorem 1. *The values of the sine and cosine functions do not exceed* 1 *in absolute value attaining the boundary values* 1 *and* -1.

Proof. Indeed, the coordinates of the point M on a unit circle do not exceed 1 in absolute value. The abscissa is equal to 1 at the right end of the horizontal diameter and to -1 at the left end, so that $1=\cos 0$ and $-1=\cos\pi$. Correspondingly, the ordinate is equal to 1 at the upper end of the vertical diameter and to -1 at the lower end, so that $1=\sin\frac{\pi}{2}$ and $-1=\sin\frac{3\pi}{2}$.

Theorem 2. $\sin^2\varphi+\cos^2\varphi=1$.

Proof. Indeed, the sum of the squares of the coordinates of a point on a circle with centre at the origin is equal to the square of the radius.

Theorem 3. $\sin\left(\varphi+\frac{\pi}{2}\right)=\cos\varphi$, $\cos\left(\varphi+\frac{\pi}{2}\right)=-\sin\varphi$.

Proof. We introduce a new system of coordinates assuming the negative direction of the original axis of ordinates to

be the positive direction of the new abscissa axis, and the original positive direction of the abscissa axis to be the positive direction of the new axis of ordinates (Fig. 4.14).

Let the radius OM be obtained from OM_0 as a result of the rotation through the angle φ. The OM is obtained from OM_0' as a result of the rotation through the angle $\varphi+\frac{\pi}{2}$, so that $\cos\left(\varphi+\frac{\pi}{2}\right)$ and $\sin\left(\varphi+\frac{\pi}{2}\right)$ are the abscissa x' and the ordinate y' of the point M with respect to the new coordinate axes. Evidently, $x'=-y$ and $y'=x$. This means that $\cos\left(\varphi+\frac{\pi}{2}\right)=-\sin\varphi$ and $\sin\left(\varphi+\frac{\pi}{2}\right)=\cos\varphi$, and this is what we wished to prove.

Theorem 4. $\sin(\varphi+2\pi)=\sin\varphi$, $\cos(\varphi+2\pi)=\cos\varphi$.

Proof. Let M be the position of a point on a unit circle corresponding to the angle φ. The addition of the angle 2π corresponds to a complete revolution of the point M about the origin and its return to the initial position. Therefore, its coordinates will return to their initial values, i.e. $\sin(\varphi+2\pi)=\sin\varphi$, $\cos(\varphi+2\pi)=\cos\varphi$.

The function f is *periodic* if there is a number $T\neq 0$ which is called a *period* and is such that $f(t+T)=f(t)$. It is clear that if $f(t+T)=f(t)$, then

$$f(t+2T)=f(t+T)=f(t),$$
$$f(t+3T)=f(t+2T+T)=f(t+2T)=f(t)$$

and so on, i.e. the equality $f(t+kT)=f(t)$ holds true for any integer k, where k may be positive as well as negative. We have proved that if T is a period, then kT is also a period for any integer k. We can prove that the periods of a continuous (and this is essential) periodic function include the least positive period. It is known as the *least period of the function.* All periods are integral multiples of the least period.

It is easy to prove that the *number* 2π *is the least period of the cosine and sine functions.* We shall prove this assertion by means of reduction ad absurdum.

We assume that the cosine function has a positive period $T'<2\pi$. Then $\cos(\varphi+T')=\cos\varphi$ for any angle φ. In particular, for $\varphi=0$ we have $\cos(0+T')=\cos 0=1$. But this means that upon a rotation through the angle T',

which is smaller than 2π, the point M_0 with abscissa 1 passes into a point whose abscissa is also 1, and this is impossible. Consequently, the cosine function has no positive period smaller than 2π. In the same way we can prove this statement for the sine function. We assume that the sine function has a positive period $T' < 2\pi$. Then $\sin(\varphi + T') = \sin\varphi$ for any angle φ. In particular, for $\varphi = \frac{\pi}{2}$ we have $\sin\left(\frac{\pi}{2} + T'\right) = \sin\frac{\pi}{2} = 1$. This means that upon a rotation through the angle T', which is smaller than 2π,

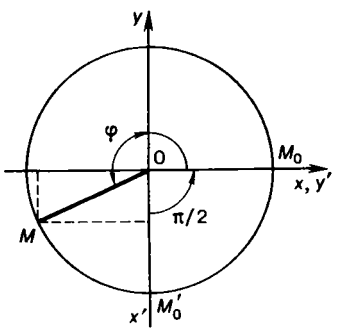


Fig. 4.14

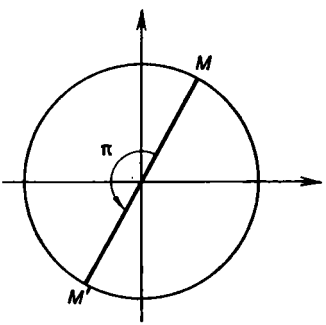


Fig. 4.15

the point with ordinate 1 passes into a point whose ordinate is also 1, and this is impossible. Consequently, 2π is the least positive period of the sine.

Theorem 5. $\cos(\varphi + \pi) = -\cos\varphi$, $\sin(\varphi + \pi) = -\sin\varphi$.

Proof. Let M be a point on a unit circle corresponding to the angle φ. Then the point M', corresponding to the angle $\varphi + \pi$, is at the end of the diameter opposite to the point M (Fig. 4.15). The coordinates of the point M' differ by signs from the coordinates of the point M. And this means that

$$\cos(\varphi + \pi) = -\cos\varphi, \quad \sin(\varphi + \pi) = -\sin\varphi.$$

Note that theorems 4 and 5 can be easily obtained from Theorem 3 by means of its formal application two and four

times. Indeed, according to Theorem 3,

$$\cos\left(\varphi+\frac{\pi}{2}\right)=-\sin\varphi,\quad \sin\left(\varphi+\frac{\pi}{2}\right)=\cos\varphi.$$

Consequently,

$$\cos(\varphi+\pi)=\cos\left(\varphi+\frac{\pi}{2}+\frac{\pi}{2}\right)=-\sin\left(\varphi+\frac{\pi}{2}\right)=-\cos\varphi,$$

$$\sin(\varphi+\pi)=\sin\left(\varphi+\frac{\pi}{2}+\frac{\pi}{2}\right)=\cos\left(\varphi+\frac{\pi}{2}\right)=-\sin\varphi.$$

Furthermore,

$$\cos\left(\varphi+\frac{3}{2}\pi\right)=\cos\left(\varphi+\pi+\frac{\pi}{2}\right)=-\sin(\varphi+\pi)=\sin\varphi,$$

$$\sin\left(\varphi+\frac{3}{2}\pi\right)=\sin\left(\varphi+\pi+\frac{\pi}{2}\right)=\cos(\varphi+\pi)=-\cos\varphi.$$

And finally,

$$\cos(\varphi+2\pi)=\cos\left(\varphi+\frac{3}{2}\pi+\frac{\pi}{2}\right)=-\sin\left(\varphi+\frac{3}{2}\pi\right)=\cos\varphi,$$

$$\sin(\varphi+2\pi)=\sin\left(\varphi+\frac{3}{2}\pi+\frac{\pi}{2}\right)=\cos\left(\varphi+\frac{3}{2}\pi\right)=\sin\varphi.$$

Theorem 6. $\cos(-\varphi)=\cos\varphi$, $\sin(-\varphi)=-\sin\varphi$.

Proof. The rotations through the angles φ and $-\varphi$ are equal in magnitude but opposite in direction. Therefore the points M and M', obtained from M_0 by means of rotations through the angles φ and $-\varphi$, lie on a unit circle and are symmetric about the abscissa axis (Fig. 4.16). Consequently, their abscissas are equal and the ordinates differ only by sign. And this means that $\cos(-\varphi)=\cos\varphi$ and $\sin(-\varphi)=-\sin\varphi$.

The function $f(x)$, whose domain of definition is symmetric about zero, is an *even function* if $f(x)=f(-x)$ for all permissible values of x and an *odd function* if $f(-x)=-f(x)$. Theorem 6 signifies that the cosine is an even function and the sine is an odd function.

Theorem 7. $\cos(\pi-\varphi)=-\cos\varphi$, $\sin(\pi-\varphi)=\sin\varphi$.

Proof.

$$\cos(\pi-\varphi)=\cos(-\varphi+\pi)=-\cos(-\varphi)=-\cos\varphi,$$

$$\sin(\pi-\varphi)=\sin(-\varphi+\pi)=-\sin(-\varphi)=\sin\varphi.$$

Theorem 8. $\cos\left(\frac{\pi}{2}-\varphi\right)=\sin\varphi$, $\sin\left(\frac{\pi}{2}-\varphi\right)=\cos\varphi$.

Proof.

$$\cos\left(\frac{\pi}{2}-\varphi\right)=\cos\left(-\varphi+\frac{\pi}{2}\right)=-\sin(-\varphi)=\sin\varphi,$$

$$\sin\left(\frac{\pi}{2}-\varphi\right)=\sin\left(-\varphi+\frac{\pi}{2}\right)=\cos(-\varphi)=\cos\varphi.$$

Note that for the acute angle φ Theorem 8 follows immediately from the definitions of the sine and cosine functions

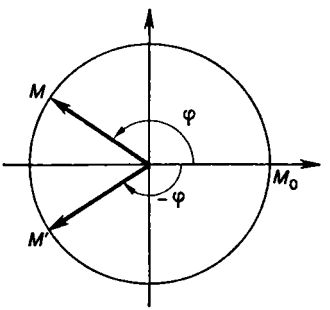


Fig. 4.16

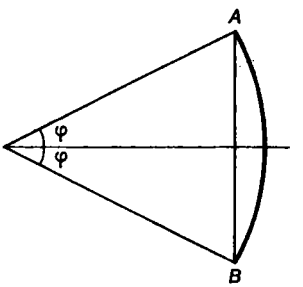


Fig. 4.17

with the aid of a right triangle with acute angle φ. Indeed, the second acute angle in this triangle is $\frac{\pi}{2}-\varphi$. The leg opposite to the angle φ is adjacent to the angle $\frac{\pi}{2}-\varphi$ and the leg adjacent to the angle φ is opposite to the angle $\frac{\pi}{2}-\varphi$. From this and from the definitions of the cosine and sine functions it follows that

$$\cos\left(\frac{\pi}{2}-\varphi\right)=\sin\varphi \quad \text{and} \quad \sin\left(\frac{\pi}{2}-\varphi\right)=\cos\varphi.$$

Theorem 9. *A chord subtending a central angle in a circle of radius 1 is equal to double the sine of the half-angle.*

Proof. Let us consider a central angle of measure 2φ, $0<2\varphi<\pi$ in a circle of radius 1. We divide it in two. Clearly, $0<\varphi<\pi/2$, i.e. φ is an acute angle. The chord AB which subtends the angle 2φ will be divided into two equal segments (Fig. 4.17) each of which is evidently equal to $\sin\varphi$. Thus $AB=2\sin\varphi$, and this is what we wished to prove.

4.5. Reducing the Values of the Sine and Cosine Functions to the Values on the Interval $0 \leqslant \varphi \leqslant \pi/4$

The values of the cosine and sine of the number φ depend only on the position of the point M, which corresponds to the angle φ, on a unit circle. The coordinate axes divide the circle into four quadrants. According as the quadrant, in which the point corresponding to the angle lies, we speak of the angles of the first quadrant, of the second quadrant, etc.

With an accuracy of an integer, which is a multiple of 2π, we can represent the angles of the first quadrant by the values from the interval $0 \leqslant \varphi \leqslant \pi/2$, the angles of the second quadrant by the values from the interval $\pi/2 \leqslant \varphi \leqslant \pi$, those of the third quadrant by the values from the interval $\pi \leqslant \varphi \leqslant 3\pi/2$ and those of the fourth quadrant by the values from the interval $3\pi/2 \leqslant \varphi \leqslant 2\pi$. Sometimes it is convenient to represent the angles of the fourth and third quadrants by the values that are smaller by 2π: by the values from the interval $-\frac{\pi}{2} \leqslant \pi \leqslant 0$ for the fourth quadrant and by the values from the interval $-\pi \leqslant \varphi \leqslant -\frac{\pi}{2}$ for the third quadrant. Boundary values are assumed to belong to both adjacent quadrants.

Theorem 5 makes it possible to reduce the calculations of the values of the cosine and sine of the angles of the third quadrant to those of the first quadrant. Theorem 7 reduces the trigonometric functions in the second quadrant to those in the first and, finally, Theorem 6 reduces the functions in the fourth quadrant to those in the first. Furthermore, if $0 \leqslant \varphi \leqslant \pi/2$ and, in addition, $\pi/4 \leqslant \varphi \leqslant \pi/2$, then $0 \leqslant \frac{\pi}{2} - \varphi \leqslant \frac{\pi}{4}$. Therefore Theorem 8 makes it possible to reduce the calculations of the cosine and sine of the angles of the first quadrant to the calculations of these functions on the interval $0 \leqslant \varphi \leqslant \pi/4$.

4.6. The Tangent and Cotangent Functions

In a right triangle with an acute angle φ the *tangent* of the acute angle φ is the ratio of the length of the leg which is opposite to the angle φ to the length of the leg adjacent

to that angle. The tangent is designated as $\tan \varphi$. The inverse ratio of the length of the leg adjacent to φ to that of the leg lying opposite to it is a *cotangent* of φ and is designated as $\cot \varphi$. It is clear that for the acute angle φ the equalities $\tan \varphi = \frac{\sin \varphi}{\cos \varphi}$ and $\cot \varphi = \frac{\cos \varphi}{\sin \varphi} = \frac{1}{\tan \varphi}$ hold true. By definition, $\tan \varphi = \frac{\sin \varphi}{\cos \varphi}$ and $\cot \varphi = \frac{\cos \varphi}{\sin \varphi}$ for any value of φ. Let us calculate the tangents and cotangents of some angles:

$$\tan 30° = \frac{\sin 30°}{\cos 30°} = \frac{1}{2} \div \frac{\sqrt{3}}{2} = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3},$$

$$\tan 45° = \frac{\sin 45°}{\cos 45°} = 1, \quad \tan 60° = \frac{\sin 60°}{\cos 60°} = \frac{\sqrt{3}}{2} \div \frac{1}{2} = \sqrt{3},$$

$$\cot 30° = \frac{1}{\tan 30°} = \sqrt{3}, \quad \cot 45° = 1, \quad \cot 60° = \frac{\sqrt{3}}{3},$$

◀ $$\tan 18° = \frac{\sqrt{5}-1}{\sqrt{10+2\sqrt{5}}} = \frac{1}{5}\sqrt{25-10\sqrt{5}},$$

$$\cot 18° = \sqrt{5+2\sqrt{5}},$$

$$\tan 72° = \sqrt{5+2\sqrt{5}}, \quad \cot 72° = \frac{1}{5}\sqrt{25-10\sqrt{5}}. \blacktriangleright$$

There is a simple geometric representation of the tangent and cotangent connected with a circle of radius 1 and centre at the origin. Namely, let $T'T$ be a tangent to a unit circle at a point M_0 of intersection of the circle with the positive direction of the abscissa axis (Fig. 4.18). Then the ordinate of the point of intersection of the tangent with the ray OM_1, corresponding to the angle φ, or with the extension of this ray (if φ refers to the second or the third quadrant) is equal to the tangent of this angle. For the acute angle φ_1 this immediately follows from the definition of the tangent and the similarity of the triangles OKM_1 and OM_0N_1. For the angle $\varphi_2 = -\varphi_1$ of the fourth quadrant the ordinate of the point N_2 and $\sin \varphi$ differ only by signs from the ordinate N_1 and $\sin \varphi_1$, and $\cos \varphi_2 = \cos \varphi_1$. Therefore the ordinate of the

point N_2 is equal to

$$-\tan\varphi_1 = \frac{-\sin\varphi_1}{\cos\varphi_1} = \frac{\sin\varphi_2}{\cos\varphi_2} = \tan\varphi_2. \text{ Furthermore,}$$

$$\tan(\varphi+\pi) = \frac{\sin(\varphi+\pi)}{\cos(\varphi+\pi)} = \frac{-\sin\varphi}{-\cos\varphi} = \tan\varphi.$$

It follows from this relation that the tangents of the angles of the third and the second quadrant are equal to the tangents

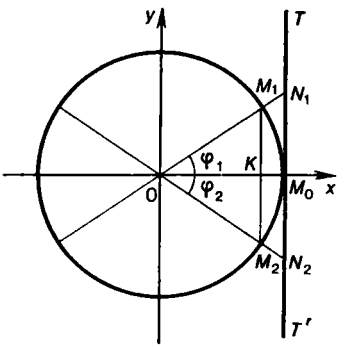


Fig. 4.18

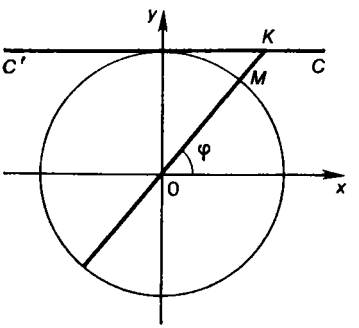


Fig. 4.19

of the angles of the first and the fourth quadrant which correspond to the extensions of the rays which define these angles.

A similar representation of the cotangent results when we introduce into consideration a tangent $C'C$ at the point of intersection of the circle with the positive direction of the axis of ordinates (Fig. 4.19). The cotangent of the angle φ is equal to the abscissa of the point K of intersection of the tangent with the ray corresponding to the angle φ or with its extension.

The tangent and cotangent functions possess some almost obvious properties.

1. For $\varphi = \pm\frac{\pi}{2} + 2k\pi$, where k is an integer, the tangent function is not defined since $\cos\left(\pm\frac{\pi}{2} + 2\pi k\right) = 0$. For φ varying from $-\pi/2$ to $\pi/2$, the values of the tangent increase and cover the entire number axis, i.e. the interval

$(-\infty, \infty)$. This follows directly from the geometric representation of the tangent (Fig. 4.18).

2. For $\varphi = \pi k$, where k is an integer, the cotangent function is not defined (since $\sin(k\pi) = 0$ for integral k). For φ varying from 0 to π, $\cot\varphi$ decreases and runs the entire length of the number axis from $+\infty$ to $-\infty$.

3. $\tan(\varphi + \pi) = \tan\varphi$, $\cot(\varphi + \pi) = \cot\varphi$.

Indeed,

$$\tan(\varphi+\pi) = \frac{\sin(\varphi+\pi)}{\cos(\varphi+\pi)} = \frac{-\sin\varphi}{-\cos\varphi} = \frac{\sin\varphi}{\cos\varphi} = \tan\varphi$$

and similarly for the cotangent. We have noted this property when we discussed the geometric interpretation of the tangent. This property means that the tangent and cotangent are periodic with period π. It is easy to see that π is the least period for $\tan\varphi$ and $\cot\varphi$.

4. $\tan\left(\frac{\pi}{2}-\varphi\right) = \cot\varphi$, $\cot\left(\frac{\pi}{2}-\varphi\right) = \tan\varphi$.

Indeed,

$$\tan\left(\frac{\pi}{2}-\varphi\right) = \frac{\sin\left(\frac{\pi}{2}-\varphi\right)}{\cos\left(\frac{\pi}{2}-\varphi\right)} = \frac{\cos\varphi}{\sin\varphi} = \cot\varphi,$$

$$\cot\left(\frac{\pi}{2}-\varphi\right) = \frac{\cos\left(\frac{\pi}{2}-\varphi\right)}{\sin\left(\frac{\pi}{2}-\varphi\right)} = \frac{\sin\varphi}{\cos\varphi} = \tan\varphi.$$

5. $\tan\left(\varphi+\frac{\pi}{2}\right) = -\cot\varphi$, $\cot\left(\varphi+\frac{\pi}{2}\right) = -\tan\varphi$.

Indeed,

$$\tan\left(\varphi+\frac{\pi}{2}\right) = \frac{\sin\left(\varphi+\frac{\pi}{2}\right)}{\cos\left(\varphi+\frac{\pi}{2}\right)} = \frac{\cos\varphi}{-\sin\varphi} = -\cot\varphi,$$

$$\cot\left(\varphi+\frac{\pi}{2}\right) = \frac{\cos\left(\varphi+\frac{\pi}{2}\right)}{\sin\left(\varphi+\frac{\pi}{2}\right)} = \frac{-\sin\varphi}{\cos\varphi} = -\tan\varphi.$$

6. The tangent and cotangent are odd functions.

Indeed,

$$\tan(-\varphi) = \frac{\sin(-\varphi)}{\cos(-\varphi)} = \frac{-\sin\varphi}{\cos\varphi} = -\tan\varphi,$$

$$\cot(-\varphi) = \frac{\cos(-\varphi)}{\sin(-\varphi)} = \frac{\cos\varphi}{-\sin\varphi} = -\cot\varphi.$$

7. $1 + \tan^2\varphi = \frac{1}{\cos^2\varphi}$, $1 + \cot^2\varphi = \frac{1}{\sin^2\varphi}$.

To derive these formulas, it is sufficient to divide both sides of the relation $\sin^2\varphi + \cos^2\varphi = 1$ by $\cos^2\varphi$ and $\sin^2\varphi$.

8. The slope of an oblique straight line $y = kx + b$ is equal to the tangent of the angle between the positive direc-

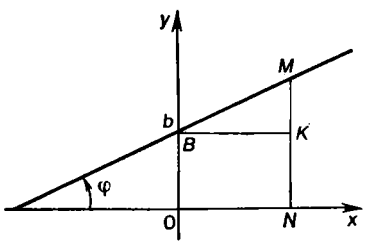


Fig. 4.20

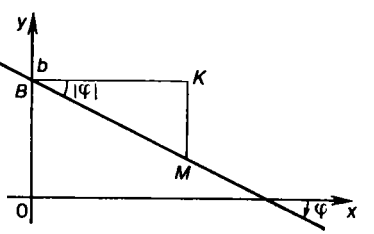


Fig. 4.21

tion of the abscissa axis and this straight line, the direction of the reckoning of this angle being taken into account.

Let the slope k be positive (Fig. 4.20). Then the straight line goes upwards when the movement is from left to right and the angle φ the straight line makes with the x-axis is positive and acute when it is reckoned in a natural way. Let $M(x, y)$ be a point on a straight line. Considering the triangle BKM, we find that $\tan\varphi = \frac{MK}{BK} = \frac{y-b}{x} = k$.

Now if the slope is negative, then the straight line goes downwards when the movement is from left to right. The angle φ the straight line makes with the x-axis is positive and obtuse or, when we reckon it in the negative direction, negative and acute. They differ by π and their tangents are similar. We shall consider a negative value of φ (Fig. 4.21). Let $M(x, y)$ be a point on a straight line. Considering the

triangle BKM, we find that

$$\tan |\varphi| = \frac{KM}{BK} = \frac{b-y}{x} = -k,$$

whence $k = -\tan |\varphi| = \tan (-|\varphi|) = \tan \varphi$.

The proof of property 8 is complete. It is a very significant property of the tangent. For the differential calculus it means that the derivative of a function at every point is equal to the tangent of the angle the x-axis makes with the tangent to the graph of the function at that point.

4.7. Expressing Trigonometric Functions in Terms of One Another

It follows from the relation $\sin^2 \varphi + \cos^2 \varphi = 1$ that $\cos \varphi = \pm\sqrt{1 - \sin^2 \varphi}$. The choice of the sign depends on the quadrant to which φ belongs: for the first and the fourth quadrant we must take the plus sign and for the second and the third quadrant, the minus sign. Furthermore,

$$\tan \varphi = \frac{\sin \varphi}{\pm\sqrt{1-\sin^2 \varphi}}$$

with the same choice of signs.

From the same relation we find that $\sin \varphi = \pm\sqrt{1 - \cos^2 \varphi}$ must be taken with the plus sign for φ belonging to the first or the second quadrant and with the minus sign for φ belonging to the third or the fourth quadrant. Furthermore,

$$\tan \varphi = \frac{\pm\sqrt{1-\cos^2 \varphi}}{\cos \varphi}$$

with the same choice of signs.

It follows from the relation $1 + \tan^2 \varphi = \frac{1}{\cos^2 \varphi}$ that $\cos^2 \varphi = \frac{1}{1+\tan^2 \varphi}$ and $\cos \varphi = \frac{1}{\pm\sqrt{1+\tan^2 \varphi}}$. Here the plus sign is taken for the first and the fourth quadrant and the minus sign for the second and the third quadrant. Furthermore, from the relation $\frac{\sin \varphi}{\cos \varphi} = \tan \varphi$ we find that

$$\sin \varphi = \tan \varphi \cos \varphi = \frac{\tan \varphi}{\pm\sqrt{1+\tan^2 \varphi}}$$

with the same choice of signs.

4.8. A Significant Limit

Theorem. $\lim\limits_{x\to 0} \dfrac{\sin x}{x} = 1$.

Proof 1. Assume that $2x_n > 0$ is a small angle which is one of the values of the variable tending to zero. We regard $2x_n$ as a central angle in a circle of radius 1 (Fig. 4.22). Then the arc that subtends this angle is $2x_n$ long and the chord that subtends the arc is $2 \sin x_n$ long. According to the main principle of differential calculus, these lengths differ by an infinitesimal α_n which tends to zero appreciably faster than $2x_n$.

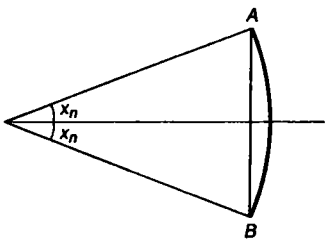


Fig. 4.22

Thus $2 \sin x_n = 2x_n - \alpha_n$ and $\dfrac{\sin x_n}{x_n} = 1 - \dfrac{\alpha_n}{2x_n}$. The second term $\dfrac{\alpha_n}{2x_n}$ tends to zero, and so $\lim \dfrac{\sin x_n}{x_n} = 1$. If $x_n < 0$, $x_n = -y_n$, then $\dfrac{\sin x_n}{x_n} = \dfrac{\sin y_n}{y_n}$. As $x_n \to 0$, $y_n \to 0$ as well, and in this case

$$\lim \frac{\sin x_n}{x_n} = \lim \frac{\sin y_n}{y_n} = 1.$$

We shall have the same limit when x_n assumes positive and negative values. Thus $\lim \dfrac{\sin x_n}{x_n} = 1$ for any sequence $x_n \to 0$ so that there is a limit of the function $\dfrac{\sin x}{x}$ as $x \to 0$ and this limit is equal to 1. And this is what we wanted to prove.

◀ *Proof 2.* However, this proof does not completely satisfy the level of strictness accepted in present-day mathematics. Here is a more rigorous proof. We recall that the length of a circle is equal to the limit of the perimeter of a regular n-gon inscribed in the circle as $n \to \infty$. We assume that the radius of the circle is 1 so that its length is 2π. The central angle that rests on a side of the regular n-gon is evidently equal to $2\pi/n$. A side of the regular n-gon, i.e.

the chord subtending this angle is $2\sin\frac{1}{2}\cdot\frac{2\pi}{n}=2\sin\frac{\pi}{n}$ long. Therefore the perimeter of the n-gon is equal to $2n\sin\frac{\pi}{n}$. Thus the relation $\lim\limits_{n\to\infty}2n\sin\frac{\pi}{n}=2\pi$ holds true, whence we have

$$\lim_{n\to\infty}\frac{\sin(\pi/n)}{\pi/n}=1.$$

Thus, for $x_n=\pi/n$, we have $\lim\limits_{x_n\to 0}\frac{\sin x_n}{x_n}=1$.

Assume now that $x_n>0$, $x_n\to 0$. Then, for every n, the number π/x_n lies between the neighbouring integers $m\leqslant\frac{\pi}{x_n}<m+1$, and, evidently, $m\to\infty$ as $x_n\to 0$. Then $\frac{\pi}{m+1}<x_n<\frac{\pi}{m}$ so that

$$\frac{\sin x_n}{x_n}<\frac{\sin(\pi/m)}{\pi/(m+1)},\quad \frac{\sin x_n}{x_n}>\frac{\sin(\pi/(m+1))}{\pi/m}.$$

The quantities $\frac{\sin(\pi/m)}{\pi/(m+1)}=\frac{m+1}{m}\cdot\frac{\sin(\pi/m)}{\pi/m}$ and $\frac{\sin(\pi/(m+1))}{\pi/m}=\frac{m}{m+1}\cdot\frac{\sin(\pi/(m+1))}{\pi/(m+1)}$ both tend to 1 as $m\to\infty$ and, consequently, $\frac{\sin x_n}{x_n}\to 1$. Furthermore, $\frac{\sin x_n}{x_n}=\frac{\sin(-x_n)}{-x_n}$ and, consequently, $\frac{\sin x_n}{x_n}=\frac{\sin|x_n|}{|x_n|}$. Therefore $\frac{\sin x_n}{x_n}\to 1$ if $x_n\to 0$ and can assume negative values as well. Thus $\lim\limits_{x\to 0}\frac{\sin x}{x}=1$ since this limiting relation is valid for any sequence x_n tending to zero. ▶

4.9. The Graphs of the Functions $y=\sin x$ and $y=\cos x$

The function $y=\sin x$ is continuous for all values of x since when a point moves continuously along a circle, its ordinate also varies continuously, i.e. for a sufficiently

small variation of x, the ordinate $\sin x$ of a point on the circle varies arbitrarily little. For x varying from 0 to $\pi/2$, i.e. when a point moves along a unit circle in the positive direction within the first quadrant, $\sin x$ evidently increases, and for $x = \pi/2$ it attains a maximum equal to 1. Furthermore, for x varying from $\pi/2$ to π, the function $\sin x$ decreases from 1 to 0. For $\pi \leqslant x \leqslant \frac{3}{2}\pi$, the function $\sin x$ keeps decreasing from 0 to -1. On the interval $\left[\frac{3}{2}\pi, \pi\right]$ the function $\sin x$ increases from -1 to 0. Then the values

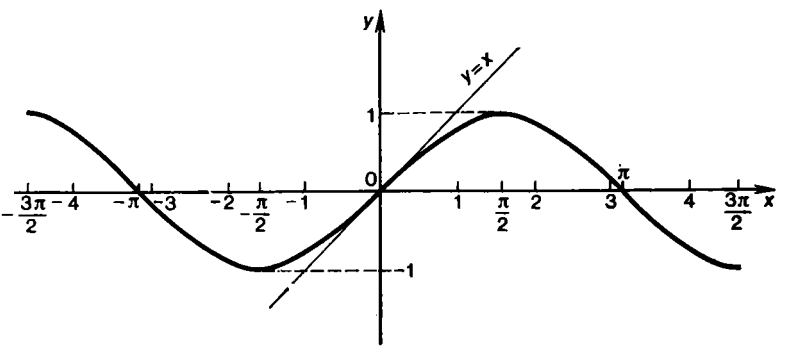


Fig. 4.23

of the function are periodically repeated, with a period 2π, irrespective of whether the point moves to the right or to the left. Thus the graph of the function $y = \sin x$ is a wavy line (Fig. 4.23) which is known as a *sine curve*.

For $0 < x \leqslant \pi/2$ the inequality $\sin x < x$ holds true since $2 \sin x$ is a chord subtending an arc $2x$ long on a unit circle. Of course, the inequality remains valid for $x > \pi/2$ as well since $\pi/2 > 1 \geqslant \sin x$. Therefore the right-hand half of the sine curve lies below the straight line $y = x$, but in the vicinity of the origin the sine curve comes close to the straight line since $\frac{\sin x}{x} \to 1$ as $x \to 0$. To be more precise, the straight line $y = x$ is a tangent to the sine curve at the point $x = 0$ and its slope 1 is equal to the derivative of the function $\sin x$ at the point $x = 0$. We can easily verify this

by means of the following simple calculation:

$$(\sin x)'_{x=0} = \lim_{\Delta x \to 0} \frac{\sin(0+\Delta x) - \sin 0}{\Delta x} = \lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} = 1.$$

Since trigonometric functions are very important for applications, tables of values of these functions have been compiled and, besides this, the values of trigonometric functions can be obtained with the use of calculators. Using the tables or calculators, we can construct sufficiently many

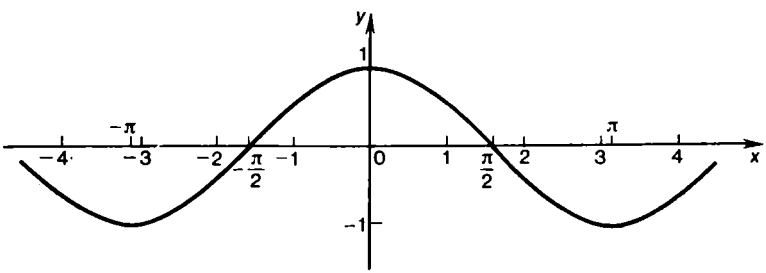


Fig. 4.24

points on the graph of $y = \sin x$. When we do this, we shall see that the sine curve bends very smoothly. We can mark many points on the graph even without a resort to tables or calculators, using only the values of $\sin x$ at the points

$$x = 0, \quad x = \pi/10 \cong 0.314, \quad x = \pi/6 \cong 0.523,$$
$$x = \pi/4 \cong 0.785, \quad x = \pi/3 \cong 1.047,$$
$$x = 2\pi/5 \cong 1.257, \quad x = \pi/2 \cong 1.571,$$

for which we calculated the values of $\sin x$ in 4.3. It is easy to construct the other points by resorting to the rules of reduction to the interval $[0, \pi/2]$.

The graph of the function $y = \cos x$ is the same sine curve but displaced to the left by $\pi/2$. Indeed, $\cos x = \sin\left(x + \frac{\pi}{2}\right)$, so that the value of the cosine at the point x is equal to the value of the sine at the point $x + \frac{\pi}{2}$ displaced to the right of the point x by $\pi/2$. The graph of the function $y = \cos x$ is shown in Fig. 4.24.

In applications we often have to deal with functions of the form $y = b \sin ax$ for $a > 0$. It is clear that the graph of such a function is a somewhat modified sine curve. Multiplication of the function by the number b leads to an extension or contraction of its graph along the axis of ordinates and to a symmetric reflection from the abscissa axis for $b < 0$. Namely, if $|b| > 1$, then the graph is extended $|b|$ times and if $|b| < 1$, then the graph is contracted $1/|b|$ times.

Multiplication of the argument of the function by the number $a > 0$ leads to a contraction or extension of its graph

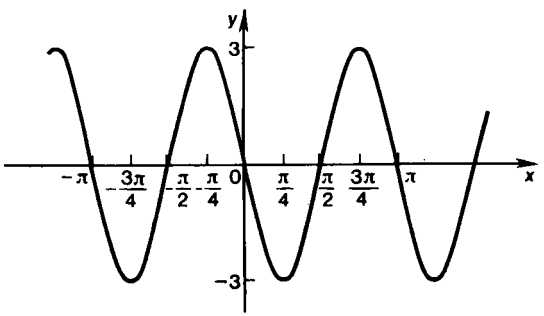


Fig. 4.25

but this time along the abscissa axis. Namely, if $a > 1$, then the graph of the function is contracted a times and if $0 < a < 1$, then the graph is extended $1/a$ times. Thus, since the graph of the function $y = \sin x$ oscillates from -1 to 1 with a period 2π, the graph of the function $y = b \sin ax$ oscillates within the limits $-|b|$ and $|b|$ with a period $T = 2\pi/a$. In Fig. 4.25, $b = -3$ and $a = 2$.

If the least period T is specified for such a function, then $a = 2\pi/T$. Therefore we can often write a function of this form as a formula $y = b \sin \frac{2\pi}{T} x$. The number $|b|$ is the *amplitude* of the function $y = b \sin ax$.

A frequently encountered function $y = b \sin (ax + c) = b \sin a (x - x_0)$ has, as its graph, or $x_0 = -c/a$, a sine curve of the same dimensions as the graph of $y = b \sin ax$ but displaced along the abscissa axis by x_0. Note that the graph of $y = \cos x = \sin\left(x + \frac{\pi}{2}\right)$ results as a special case for $a = b = 1$, $x_0 = -\pi/2$.

4.10. The Graphs of the Functions $y = \tan x$ and $y = \cot x$

As we have seen, from the geometric representation of the tangent (Fig. 4.18) it follows immediately that $y = \tan x$ increases from $-\infty$ to $+\infty$ as x varies on the interval $(-\pi/2, \pi/2)$. For $x = -\pi/2$ and $x = \pi/2$ the tangent has no sense. Sometimes we write $\tan \pi/2 = \infty$ but this is a conventional notation of the fact that $\tan x \to \infty$ as $x \to \pi/2$

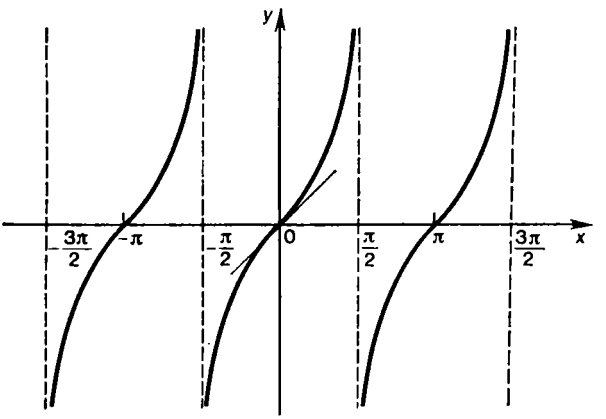


Fig. 4.26

(in more detail, $\tan x \to +\infty$ as $x \to \pi/2$ and $x < \pi/2$ and $\tan x \to -\infty$ as $x \to \frac{\pi}{2}$ and $x > \frac{\pi}{2}$). On $(-\pi/2, \pi/2)$ the graph of the function $y = \tan x$ is shaped as a rising curve. The number π is a period of $\tan x$ and therefore the curve representing $\tan x$ on $(-\pi/2, \pi/2)$ is repeated periodically many times with a period π (Fig. 4.26). The graph of the function $y = \tan x$ is sometimes called a *tangent curve.* It cuts the abscissa axis at an angle of 45°. Indeed, for x close to 0, $\cos x$ approaches 1 and $\sin x \cong x$, so that

$$\tan x = \frac{\sin x}{\cos x} \cong \frac{x}{1} = x.$$

(To be more precise, we must say that for $x = 0$ the value of the derivative of $\tan x$ is 1. But this is almost obvious:

$$(\tan x)'_{x=0} = \lim_{\Delta x \to 0} \frac{\tan(0+\Delta x) - \tan 0}{\Delta x} = \lim_{\Delta x \to 0} \frac{\tan \Delta x}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{\sin \Delta x}{\Delta x} \cdot \frac{1}{\cos \Delta x} = 1.)$$

We have seen (Fig. 4.19) that as x varies on the interval $(0, \pi)$, the function $y = \cot x$ decreases from $+\infty$ to $-\infty$

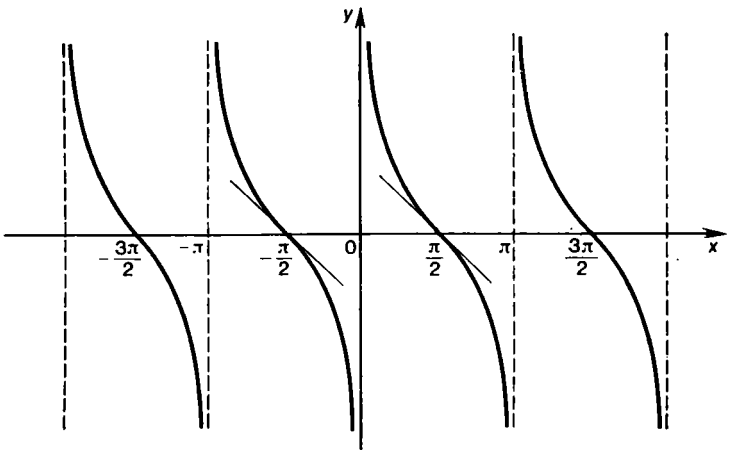


Fig. 4.27

and passes through 0 for $x = \pi/2$. Then the graph of $y = \cot x$ is repeated periodically with a period π (Fig. 4.27). The graph cuts the abscissa axis at an angle of $-45°$. Indeed,

$$(\cot x)'_{x=\frac{\pi}{2}} = \lim_{\Delta x \to 0} \frac{\cot\left(\frac{\pi}{2} + \Delta x\right) - \cot\frac{\pi}{2}}{\Delta x}$$

$$= \lim_{\Delta x \to 0} \frac{-\tan \Delta x}{\Delta x} = -1.$$

EXERCISES TO CHAPTER 4

1. Represent in the coordinate plane the result of a rotation of the positive half-axis of abscissas through a specified angle $\alpha = p°$ (if $|p| > 360$, then first isolate complete rotations by dividing

$|p|$ by 360 with a remainder):

1. 800°, −230°, 2000°, −1400°, 3960°,
2. 750°, −200°, 2380°, −1740°, 4680°.

Sample solution of the problem. Let $\alpha = 1400°$, and then

$$1400° = 360° \cdot 3 + 320°.$$

To make the construction easier, the angle of 320° must be represented as $320° = 180° + 140°$ or $320° = 360° - 40°$. The answer to the problem is given in Fig. 4.28.

2. We say that the *results of the rotations through the angles* α *and* β *are the same* if the positive half-axis of abscissas occupies the same

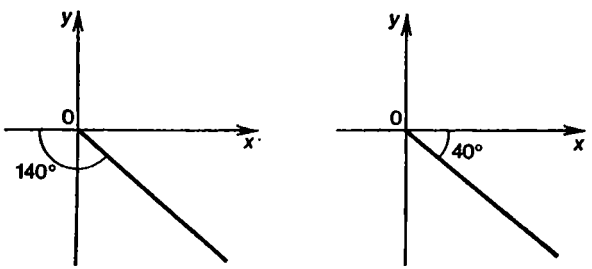


Fig. 4.28

position upon the rotations through the angles α and β. Which of the rotation angles coincide?

1. 450°, 1040°, 1400°, −990°, −2220°, −1720°,
2. 630°, 2430°, 1140°, −2610°, 660°, −2820°.

3. Find the angle $\alpha = p°$, $|p| \leqslant 180$, the result of the rotation through which coincides with the result of the rotation through the given angle β:

1. $\beta = 400°$, −840°, 1000°, 1410°.
2. $\beta = -510°$, 800°, −920°, 2015°.

4. Show that the result of the rotation through any angle coincides with the result of the rotation through an angle $\alpha = p°$, where $|p| \leqslant 180$.

5. It is clear that if the rotation angles measured in degrees differ by a value which is a multiple of 360, then the results of the rotations coincide. Prove the converse statement.

6. Write the set of all angles upon a rotation through which the positive half-axis of abscissas occupies the position marked by the ray OM in Fig. 4.29.

Sample solution of the problem. One of the angles, upon a rotation through which the x-axis occupies this position, is equal to 120° (Fig. 4.30). The other angles differ from 120° by a value which is a

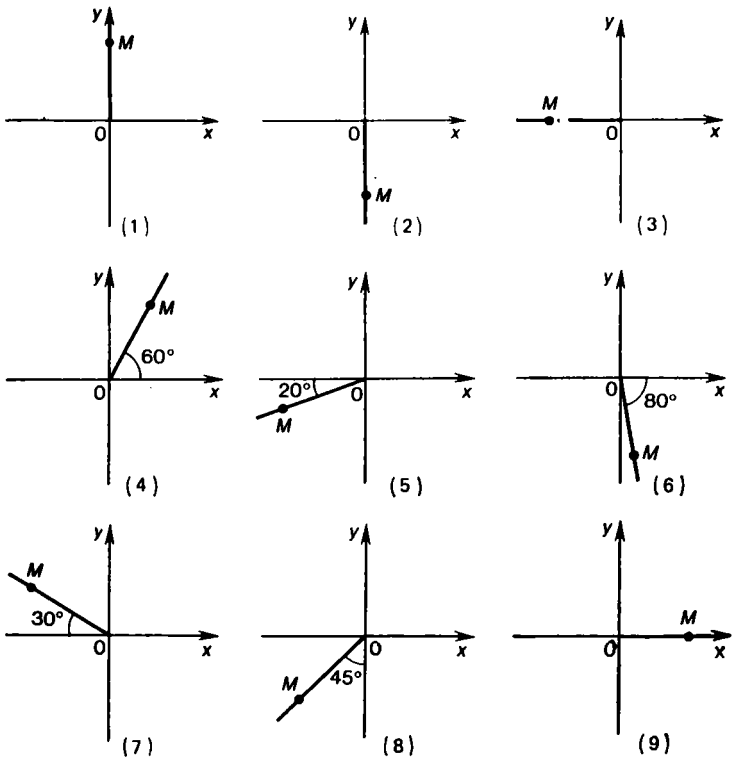


Fig. 4.29.

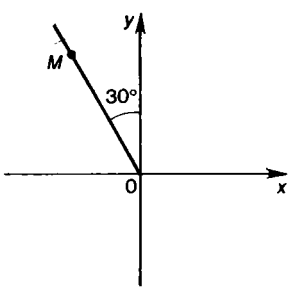


Fig. 4.30

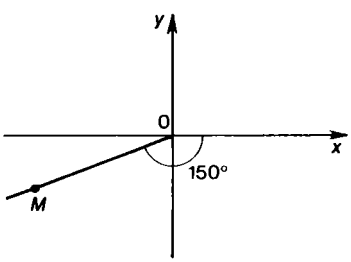


Fig. 4.31

multiple of 360. We can write this set of angles as $120° + 360° \cdot n$, $n \in \mathbb{Z}$ ($\mathbb{Z}$ is the set of integers).*

7. Represent in the coordinate plane the result of the rotation of the positive half-axis of abscissas through angles belonging to the given set and name some angles of each set:

1. $-30° + 360° \cdot n$, $n \in \mathbb{Z}$. 2. $150° + 360° \cdot n$, $n \in \mathbb{Z}$.
3. $180° + 360° \cdot n$, $n \in \mathbb{Z}$. 4. $-180° + 360° \cdot n$, $n \in \mathbb{Z}$.
5. $-120° + 360° \cdot n$, $n \in \mathbb{Z}$. 6. $45° + 360° \cdot n$, $n \in \mathbb{Z}$.

Sample solution of the problem. Any two angles of the set $-150° + 360° \cdot n$, $n \in \mathbb{Z}$, differ by a value which is a multiple of 360°. Consequently, the results of rotation through these angles coincide. Therefore it is sufficient to isolate any angle from the set and depict the corresponding rotation. We have $n = 0$ and obtain an angle of $-150°$. The answer is given in Fig. 4.31. To name some specific angles of the given set, we must assign some integral values to the parameter n. For instance, for $n = 1$, $\alpha = 210°$, for $n = -2$, $\alpha = -870°$, and so on.

8. Convert each of the given degree measures into a radian measure:

1. 30°, −45°, 150°, 18°, −1000°.
2. 60°, −135°, 270°, 27°, −2000°.

9. Convert each of the given radian measures into a degree measure:

1. $\frac{\pi}{12}$, $-\frac{\pi}{2}$, $-\frac{7}{6}\pi$, $-\pi$, $\frac{2}{3}\pi$, -26π,
2. $\frac{\pi}{18}$, $-\frac{3}{2}\pi$, π, $-\frac{4}{9}\pi$, $\frac{11}{6}\pi$, -18π.

10. Associate every number x with a point of a unit circle through which the positive half-axis of abscissas passes as a result of rotation through an angle of x radians. Thus the number $\frac{\pi}{2}$, for example, will be associated with a point with coordinates (0, 1).

On a unit circle mark points corresponding to the given numbers (rotation angles):

1. $-\frac{\pi}{2}$, $\frac{3}{4}\pi$, $-\pi$, $\frac{\pi}{3}$, $-14\frac{5}{6}\pi$.
2. $-\frac{\pi}{4}$, $\frac{\pi}{6}$, $-\frac{2}{3}\pi$, π, $-8\frac{1}{3}\pi$.

11. On a unit circle mark (by sight) an approximate position of the points which correspond to the following numbers:
1. 1, 2, 3, 4, 5, 6, 2. −1, −2, −3, −4, −5, −6.

---

* The sign $\in$ stands for the word "belongs".

12. Plot an approximate position of each of the following numbers on the coordinate axis with integers marked on it (see Fig. 4.32):

$$\pm\frac{\pi}{6},\ \pm\frac{\pi}{4},\ \pm\frac{\pi}{3},\ \pm\frac{\pi}{2},\ \pm\pi,\ \pm\frac{3}{2}\pi,\ \pm 2\pi.$$

13. On a unit circle mark points corresponding to each of the numbers of the specified set:

1. $\frac{\pi}{4}+2\pi k,\ k\in\mathbb{Z}$. 2. $-\frac{\pi}{6}+2\pi k,\ k\in\mathbb{Z}$.

3. $-\frac{2}{3}\pi+2\pi k,\ k\in\mathbb{Z}$. 4. $\frac{5}{6}\pi+2\pi k,\ k\in\mathbb{Z}$.

5. $\pi+2\pi k,\ k\in\mathbb{Z}$. 6. $-\pi+2\pi k,\ k\in\mathbb{Z}$.

Sample solution of the problem. We are given a set of numbers $-\frac{\pi}{3}+2\pi k,\ k\in\mathbb{Z}$. Any two numbers of this set differ by a value which is a multiple of 2π. Hence, the results of the rotations through

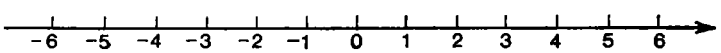


Fig. 4.32

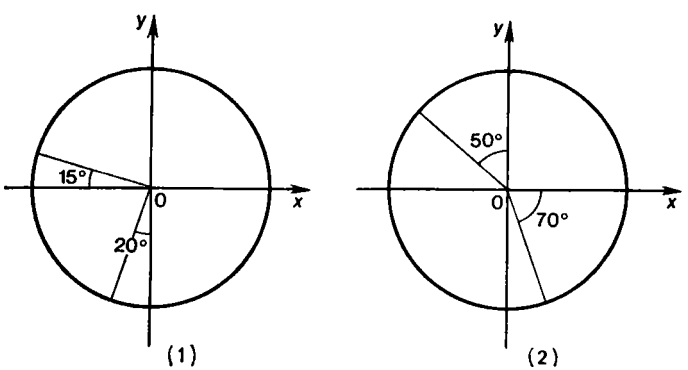


Fig. 4.33

these angles coincide, i.e. all the numbers are associated with the same point. It is sufficient to find the point corresponding to one of the numbers. For $k=0$ we have $-\pi/3$.

14. For each of the points marked on a unit circle write a set of numbers to which it corresponds (Fig. 4.33).

15. On a unit circle mark points corresponding to the numbers of the following sets:

1. $\frac{\pi}{2}+\pi n, \quad n \in \mathbb{Z}$.
2. $-\frac{\pi}{4}+\pi n, \quad n \in \mathbb{Z}$.
3. $-\pi+\pi n, \quad n \in \mathbb{Z}$.
4. $\pi n, \quad n \in \mathbb{Z}$.
5. $-\frac{\pi}{3}+\frac{\pi}{4} n, \quad n \in \mathbb{Z}$.
6. $\frac{\pi}{3}+\frac{\pi}{4} n, \quad n \in \mathbb{Z}$.

16. There are N equispaced points on a unit circle. One of them is associated with a number α. Write the set of all numbers each of which is associated with one of the points.

17. Derive a formula for the length of arc expressed in terms of the radius of a circle and the measure of the central angle corresponding to it in radians.

18. Find the angle through which the minute hand rotates in 10, 15, 30, 45 and 60 minutes. Express the measure of the angle in degrees and in radians.

19. Find the angle (understanding it in a generalized sense, i.e. taking into account the number of complete revolutions) through which

(1) the minute hand rotates in 600 min, in 1300 min, in 1.4 h,

(2) the hour hand rotates in 1, 30, 240 min, in 12, 24, 2340 h.

Express these angles in degrees and radians.

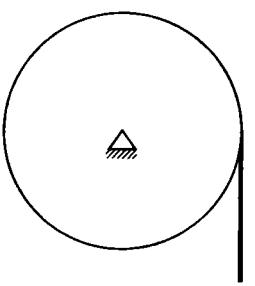


Fig. 4.34

20. Compose a function which would express the relation between the angle of rotation of the minute hand of a clock as a function of time, understanding the angle in a generalized sense and taking 1 minute as a unit of measurement. Express the angle in degrees and radians.

21. What path does the minute hand of a hand watch traverse during an hour, 24 hours, a year, if the distance from the tip of the hand to the axis is 1 cm?

22. A thin inelastic thread is wound on a disk with radius of 1 m, the end of the thread hanging down from the disk as shown in Fig. 4.34. The disk rotates about its axis. The angle of rotation varies according to the law $\varphi(t) = 20t - t^2$, where φ is in radians and t in seconds. The wound and the hanging down parts of the thread are assumed to be infinitely long. The counterclockwise direction of rotation is positive. Find

(1) the length of the thread wound on the disk or wound off the disk up to the moment

$$t = 5\,\text{s},\ 15\,\text{s},\ 20\,\text{s},\ 30\,\text{s},$$

(2) the time during which the thread was being wound on the disk and the maximum length of the part of the thread wound on the disk,

(3) the moment, reckoned from the start of the motion, at which the thread returned to the initial position,

(4) the path a point have taken on the circumference of the disk traversed during the time

$$t = 5 \text{ s}, 15 \text{ s}, 20 \text{ s}, 30 \text{ s},$$

(5) the changes that will occur in the answers to the preceding questions if the radius is equal to r m.

23. Draw a unit circle in the xOy system of coordinates. Using this circle, compile a table of sines and cosines of the following angles (numbers):

$$0, \pm\frac{\pi}{2}, \pm\pi, \pm\frac{3}{2}\pi, \pm 2\pi, \pm\frac{5}{2}\pi, \pm 3\pi, \pm\frac{7}{2}\pi, \pm 4\pi.$$

It is not necessary to memorise the table, you must know how to find the sine and cosine of these angles (numbers) orally.

24. Carry out an assignment similar to the preceding one for the same angles written in degree measure.

25. Compile a table of values of the sines and cosines of the following angles (numbers): $0, \pi/6, \pi/4, \pi/3, \pi/2$.

Pay attention to the fact that the values of the sines of these angles form an interesting progression: $\sqrt{0}/2, \sqrt{1}/2, \sqrt{2}/2, \sqrt{3}/2, \sqrt{4}/2$. Analyse the changes in the sine and cosine with an increase in the angle and compare the values of the sine and cosine of different angles. It pays to memorise this table since its data are often encountered in school problems in mathematics and physics.

26. Proceeding from the definition of trigonometric functions, the properties of rotation through an arbitrary angle and using the table from the preceding problem, find the sines and cosines of the following angles (numbers):

1. $\frac{\pi}{2}+2\pi k, \quad k \in \mathbb{Z}.$
2. $\pi+2\pi k, \quad k \in \mathbb{Z}.$
3. $-\pi+2\pi k, \quad k \in \mathbb{Z}.$
4. $-\frac{\pi}{2}+2\pi k, \quad k \in \mathbb{Z}.$
5. $2\pi k, \quad k \in \mathbb{Z}.$
6. $\frac{3}{2}\pi+2\pi k, \quad k \in \mathbb{Z}.$
7. $\frac{5}{6}\pi+2\pi k, \quad k \in \mathbb{Z}.$
8. $-\frac{\pi}{3}+2\pi k, \quad k \in \mathbb{Z}.$

27. Using a sheet of millimetre graph paper, construct a unit circle taking 5 cm as a unit of measurement. With the aid of this circle and a protractor, find the sines and cosines of the angles given below with the maximum possible accuracy. Think of the way to verify your answers with the use of four-digit tables of values of sines and cosines of acute angles. (If the accuracy in construction is high enough, the error will not exceed 0.04.)

1. $\frac{\pi}{9}+2\pi n, \quad n \in \mathbb{Z}.$
2. $\frac{5}{9}\pi+2\pi n, \quad n \in \mathbb{Z}.$
3. $1+2\pi n, \quad n \in \mathbb{Z}.$
4. $2.5+2\pi n, \quad n \in \mathbb{Z}.$

5. $-\frac{5}{8}\pi+2\pi n$, $n\in\mathbb{Z}$. 6. $-\frac{\pi}{8}+2\pi n$, $n\in\mathbb{Z}$.

7. $-2.5+2\pi n$, $n\in\mathbb{Z}$. 8. $-1+2\pi n$, $n\in\mathbb{Z}$.

28. Find all numbers satisfying each of the following equations:

1. $\sin x = 1$. 2. $\sin x = -1$. 3. $\cos x = -1$.
4. $\cos x = 1$. 5. $\sin x = 0$. 6. $\cos x = 0$.
7. $\sin x = 1/2$. 8. $\sin x = \sqrt{3}/2$. 9. $\cos x = 1/2$.
10. $\cos x = \sqrt{3}/2$. 11. $\sin x = -\sqrt{3}/2$. 12. $\cos x = -1/2$.
13. $\cos x = -\sqrt{3}/2$. 14. $\sin x = -1/2$.
15. $\sin x = 1/\sqrt{2}$. 16. $\cos x = 1/\sqrt{2}$.

29. Using a unit circle constructed on the millimetre graph paper and a protractor, find an approximate solution of each of the following equations and write the answer in radian measure. Check the answer using four-digit tables of values of sines and cosines of acute angles and the table for conversion of a degree measure into a radian measure.

1. $\sin x = 0.32$. 2. $\sin x = 0.74$ 3. $\cos x = 0.26$.
4. $\cos x = 0.66$. 5. $\sin x = -0.20$. 6. $\sin x = -0.82$.
7. $\cos x = -0.44$. 8. $\cos x = -0.80$.

30. Find the numbers from the interval $[0, 2\pi]$ which satisfy each of the following equations:

1. $2\sin^2 x + 3\sin x - 2 = 0$.
2. $2\cos^2 x - (2-\sqrt{3})\cos x - \sqrt{3} = 0$.
3. $2\cos^2 x + (4+\sqrt{2})\cos x + 2\sqrt{2} = 0$.
4. $2\sin^2 x - (2+\sqrt{2})\sin x + \sqrt{2} = 0$.

31. Solve the following equations:

1. $4\sin x = 4 - \cos^2 x$. 2. $3 + 3\sin x = \cos^2 x$.
3. $\cos^2 x - \sin^2 x = \sin x$. 4. $1 - 2\cos x = \sin^2 x - \cos^2 x$.

32. Find the solution of each of the following equations that would satisfy the inequality $0 \leqslant x \leqslant \pi$:

1. $\sin\left(2x+\frac{\pi}{3}\right) = -\frac{\sqrt{3}}{2}$. 2. $\cos\left(2x+\frac{\pi}{4}\right) = -\frac{\sqrt{2}}{2}$.
3. $\cos\left(\frac{1}{2}x-\frac{\pi}{6}\right) = \frac{\sqrt{3}}{2}$. 4. $\sin\left(\frac{x}{2}-\frac{\pi}{6}\right) = \frac{1}{2}$.
5. $\sin(3x-\pi) = 0$. 6. $\cos(3x-\pi) = -1$.

33. Solve the following equations:

1. $\sin\left(x+\frac{\pi}{2}\right)+\cos x = -1$. 2. $\sin x - \cos\left(x+\frac{\pi}{2}\right) = 1$.
3. $\sin\left(x-\frac{\pi}{3}\right)-\cos\left(x+\frac{\pi}{6}\right) = \sqrt{2}$.

4. $\cos\left(x-\frac{\pi}{6}\right)-\sin\left(x+\frac{\pi}{3}\right)=-\sqrt{2}$.

5. $\cos\left(x-\frac{\pi}{4}\right)+\sin\left(x+\frac{\pi}{4}\right)=-\sqrt{3}$.

6. $\sin\left(x-\frac{\pi}{4}\right)-\cos\left(x+\frac{\pi}{4}\right)=\sqrt{3}$.

34. A point M moves along the circumference of a unit circle with centre at the origin. The position of the point varies according to the law $\alpha=\omega t+\alpha_0$, where t is time in seconds and α is the angle in radians between the ray OM and the half-line Ox understood in the generalized sense.

1. Determine the physical meaning and the dimensionality of the parameters ω and α_0.

2. What is the minimum time interval in which the point M passes one and the same fixed position? You can get the answer proceeding from two considerations:

(a) the successive values of the rotation angles for which the point occupies the same position differ by 2π,

(b) the minimum path the point must traverse in order to return to a certain position on the circle is equal to the length of the circle, i.e. 2π, and the linear velocity of motion along the circle is equal to the angular velocity multiplied by the radius of the circle.

35. Solve problem 34 for the following specific laws of motion:

1. $\alpha=2t+\frac{\pi}{4}$. 2. $\alpha=\frac{1}{2}t-\frac{\pi}{3}$. 3. $\alpha=-\frac{2}{5}t$.

4. $\alpha=-\frac{4}{3}t$. 5. $\alpha=\frac{\pi}{2}t$. 6. $\alpha=2\pi t$.

36. Show that each of the following functions is periodic and find its least positive period:

1. $f(x)=\sin\left(2x+\frac{\pi}{3}\right)$. 2. $g(x)=\cos\left(4x-\frac{\pi}{4}\right)$.

3. $f(t)=\cos\frac{2}{5}t$. 4. $\varphi(t)=\sin\frac{t}{2}$.

5. $\varphi(t)=\sin(t-1)$. 6. $h(t)=\cos(3-t)$.

37. The value of the function $y=\sin x$ is $\sqrt{3}/2$ for $x=\pi/3$. Find the positive and the negative value of the argument the closest to $\pi/3$ and for which the function $\sin x$ is equal to $\sqrt{3}/2$. Find out whether the answer to the problem contradicts the fact that the least positive period of $\sin x$ is 2π.

38. Prove that π is a period of the functions $y=|\sin x|$ and $y=|\cos x|$.

39. Find out which of the following functions are even and which are odd and which do not possess either of the properties:

1. $f(x)=x^2$, $g(x)=x^3$, $h(x)=(x-1)^2$.
2. $f(x)=\dfrac{x}{x^2+1}$, $g(x)=\sqrt{x^2-3}$, $h(x)=\dfrac{x}{\sqrt{x+7}}$.
3. $f(x)=\sin x+\cos x$, $g(x)=\sin x\cos x$, $h(x)=\cos 2x\cdot\cos x$.
4. $f(x)=\sin x+\sin 2x$, $g(x)=\cos x-\cos 2x$, $h(x)=\cos\left(x-\dfrac{\pi}{4}\right)$.

40. Show that any function of the form $y=f(x)+f(-x)$ is even and any function of the form $y=f(x)-f(-x)$ is odd.

41. Prove that any function defined on $(-a, a)$ is the sum of an even and an odd function.

42. Prove that the graph of an even function is symmetric with respect to the axis of ordinates and the graph of an odd function is symmetric with respect to the origin.

43. Let $f(x)$ be an odd function defined on the interval $(-a, a)$. What is $f(0)$ equal to?

44. Reduce the following trigonometric functions to the trigonometric functions of acute angles; with the aid of appropriate tables find their values with an accuracy of 0.01:

1. $\sin 142°$, $\cos 165°$, $\sin(-170°)$, $\cos(-115°)$.
2. $\sin 222°$, $\cos 254°$, $\sin(-258°)$, $\cos(-193°)$.
3. $\sin 314°$, $\cos 351°$, $\sin(-348°)$, $\cos(-326°)$.
4. $\sin 854°$, $\cos 1196°$, $\sin(-1247°)$, $\cos(-861°)$.
5. $\sin 0.8\pi$, $\cos 0.6\pi$, $\sin(-0.65\pi)$, $\cos(-0.9\pi)$.
6. $\sin\left(1\dfrac{2}{7}\pi\right)$, $\cos(1.3\pi)$, $\sin\left(-1\dfrac{4}{9}\pi\right)$, $\cos\left(-1\dfrac{1}{5}\pi\right)$.
7. $\sin(1.9\pi)$, $\cos(1.6\pi)$, $\sin(-1.75\pi)$, $\cos(-1.85\pi)$.
8. $\sin 14.2\pi$, $\cos 24.3\pi$, $\sin(-13.7\pi)$, $\cos(-7.8\pi)$.

45. Prove the validity of the following equalities:

1. $\sin(\pi-\alpha)+\cos\left(\dfrac{\pi}{2}+\alpha\right)+\sin\left(\dfrac{3}{2}\pi+\alpha\right)+\cos(-\alpha)=0$.
2. $\cos(\pi-\alpha)+\sin\left(\dfrac{3}{2}\pi+\alpha\right)-\cos(\pi+\alpha)+\sin\left(\dfrac{\pi}{2}+\alpha\right)=0$.
3. $\cos^2(\pi-\alpha)+\cos^2\left(\dfrac{3}{2}\pi+\alpha\right)+\sin^2(\pi+\alpha)+\sin^2\left(\dfrac{\pi}{2}+\alpha\right)=2$.
4. $\sin^2\left(\dfrac{3}{2}\pi-\alpha\right)+\cos^2\left(\dfrac{\pi}{2}+\alpha\right)+\sin^2(\pi-\alpha)+\cos^2(\pi+\alpha)=2$.
5. $\sin(\alpha-\pi)+\sin\left(\dfrac{\pi}{4}-\alpha\right)-\cos\left(\alpha-\dfrac{3}{2}\pi\right)-\cos\left(\alpha+\dfrac{\pi}{4}\right)=0$.
6. $\cos\left(\alpha-\dfrac{\pi}{3}\right)+\sin\left(\alpha-\dfrac{3}{2}\pi\right)+\cos(\alpha-\pi)-\sin\left(\alpha+\dfrac{\pi}{6}\right)=0$.

46. Reduce the following trigonometric functions to the values on the interval $[0, \pi/4]$ $(0° \leqslant \varphi \leqslant 45°)$ in the corresponding cases):

1. $\sin \frac{\pi}{3}$, $\cos \frac{\pi}{3}$, $\sin \frac{2}{3}\pi$, $\cos \frac{2}{3}\pi$.

2. $\sin \frac{3}{7}\pi$, $\cos \frac{4}{9}\pi$, $\sin \frac{7}{8}\pi$, $\cos 0.9\pi$.

3. $\sin (-230°)$, $\cos (-200°)$, $\sin 600°$, $\cos 970°$.

4. $\sin (1000°)$, $\cos (-670°)$, $\sin (210°)$, $\cos 250°$.

5. $\sin 1$, $\cos 2$, $\sin (-2)$, $\cos (-1)$.

6. $\sin 3$, $\cos (-3)$, $\sin (-6)$, $\cos 6$.

47. Set up an expression for the perimeter P_n of a regular n-gon inscribed in a circle of unit radius. Consider the sequence P_n $(n \geqslant 3)$ of the perimeters of regular polygons inscribed in the circle writing out the approximate values (with the accuracy of 0.01) of its several first terms. Find the term of the sequence which differs from the length of the circle by a value smaller than 0.1. Find the error resulting from the replacement of the length of the circle by the perimeter of 60-gon.

48. Take a sheet of millimetre graph paper not smaller than 15×20 cm in size and construct on it the first quadrant of a unit circle assuming 10 cm to be a scale unit. Using the geometric description of the tangent, compile a table of tangents of the angles from 0° to 60° with a stepsize of 10°. Convert the values of the angles into the radian measure. Compare the values obtained for the tangent with the tabular values and estimate the relative error of the measurements.

49. Using the drawing made for the preceding problem, compare the values of $\tan x$ and $\sin x$ when x varies from 0 to $\pi/18$. Compare the tabular values of $\tan x$ and $\sin x$ for the angles from the interval of 0 to $\pi/18$.

50. Take a sheet of millimetre graph paper not smaller than 5×25 cm in size and construct on it a unit circle assuming 1 cm to be a scale unit. Using the geometric description of the tangent, compile a table of tangents of angles beginning with the angle of 60° with the stepsize of 3°. Compare the results obtained with the tabular ones and estimate the relative error of measurements. Using the table, compare the values of the functions $y = \tan x$ and $y = 1/\cos x$ in the vicinity of the point $x = \pi/2$.

51. Explain the character of the variation of $\tan x$ in the vicinity of the points $x = 0$ and $x = \pi/2$, proceeding from the definition of $\tan x$ and the character of variation of the functions $\sin x$ and $\cos x$ in the vicinity of these points.

52. Using the geometric description, study the behaviour of the function $y = \cot x$ on the interval from 0 to π. Compare the values of $\cot x$ and $\cos x$ in the vicinity of the point $x = \pi/2$ and the values of $\cot x$ and $1/\sin x$ in the vicinity of the points $x = 0$ and $x = \pi$.

53. Take a sheet of millimetre graph paper and construct a unit circle and tangents to it at the points (1,0) and (0, 1), assuming 5 cm

to be a scale unit. Find approximate solutions of the equations. Write the general solution and, separately, the solution belonging to the interval $(-\pi/2, \pi/2)$ for the tangents and to the interval $(0, \pi)$ for the cotangents.

1. $\tan x = 0.7$, 2. $\tan x = -0.5$. 3. $\cot x = -1.4$,
4. $\cot x = 1.8$. 5. $\tan x = 0.2$. 6. $\cot x = 0.9$.
7. $\tan x = -1.2$. 8. $\tan x = 1.9$.

54. Find the smallest positive period of each of the following functions:

1. $\tan\left(2x-\frac{\pi}{4}\right)$. 2. $\cot\left(\frac{x}{2}+\frac{\pi}{3}\right)$.

3. $\cot\left(\frac{1}{3}x-\frac{\pi}{4}\right)$. 4. $\tan\left(3x+\frac{\pi}{6}\right)$.

55. Reduce the following functions to the values on the interval $(0, \pi/4)$ ($(0°, 45°)$ in the corresponding cases):

1. $\tan\frac{7}{6}\pi$. 2. $\cot 0.9\pi$. 3. $\cot 600°$. 4. $\tan 980°$.

5. $\tan 3.7\pi$. 6. $\cot 5.8\pi$. 7. $\cot 3$. 8. $\tan 6$.

56. The graph of the function $y = f(x)$ is a semicircle with centre at the origin and radius equal to 5, lying in the upper half-plane. Carefully carrying out the necessary constructions and measurements on a sheet of millimetre graph paper, calculate $f'(-4.5)$, $f'(-4)$, $f'(-3)$, $f'(-2)$, $f'(-1)$, $f'(0)$. Using the property of tangents, find the values of the derivative at the points $x = 1, 2, 3, 4, 4.5$. Write the equation of the function $y = f(x)$ and calculate the value of the derivative $f'(x)$ at these points having differentiated $f(x)$. Compare the answers (if you have carefully constructed the graph and the tangents, the error must not exceed 0.01). How will the derivative of the function $f(x)$ change as $x \to -5$, $x \to 5$?

57. Knowing the value of one of the trigonometric functions, find the values of the other functions for the same value of the arguments:

1. $\sin x = 0.3$, $x \in \left(\frac{\pi}{2}, \pi\right)$.

2. $\cos x = -0.4$, $x \in \left(-\frac{\pi}{2}, \pi\right)$.

3. $\tan x = 1.2$, $x \in \left(-\pi, -\frac{\pi}{2}\right)$.

4. $\cot x = -0.8$, $x \in \left(-\frac{\pi}{2}, 0\right)$.

58. Using the appropriate tables or a calculator, find the angles beginning with which the value of $\sin x$ can be replaced by the angle itself, expressed in radians, if the error in the definition of the sine must not exceed: (a) 0.01, (b) 0.001, (c) 0.0001.

59. Show that $\cos x \cong \frac{\pi}{2} - x$, for x sufficiently close to $\frac{\pi}{2}$. Substantiate this approximate equality by the reference to the respective formula, verify it by using a table or a calculator.

60. Take a sheet of millimetre graph paper and construct the graphs of the functions $y = \sin x$ and $y = \cos x$, assuming 3 cm to be a scale unit used on both sides of the axis. First mark the approximate positions of the points corresponding to the numbers $\pm \frac{\pi}{10}$, $\pm \frac{\pi}{6}$, $\pm \frac{\pi}{4}$, $\pm \frac{\pi}{3}$, $\pm \frac{\pi}{2}$, $\pm \pi$ on the abscissa axis. Construct the graphs on the interval $[-\pi, \pi]$. When doing the construction, take account of the fact that on $[\pi/2, \pi]$ the graph of $\sin x$ is symmetric with respect to the graph of $\sin x$ on $[0, \pi/2]$ about the straight line $x = \pi/2$, and on $[-\pi, 0]$ it is symmetric with respect to the graph of $\sin x$ on $[0, \pi]$ about the origin. (Substantiate these statements using the properties of $\sin x$.) Before constructing the graph of $\sin x$ on $[0, \pi/2]$, draw a straight line $y = x$. Pay attention to the fact that within the limits of the accuracy of the drawing the graph of $\sin x$ on the interval $[0, \pi/6]$ practically merges with the straight line $y = x$.

61. Carrying out assignment 60, you have probably noticed that with the customary choice of the scale unit laid off along the abscissa axis (the unit is an integral number of divisions of the axis, in this particular case the unit comprises 30 small divisions), it is difficult to find the points corresponding to the "good" values of the argument of $\sin x$ and $\cos x$ such, for instance, as $\pi/6$, $\pi/3$, $\pi/2$, π and so on, and therefore when constructing the graphs of trigonometric functions it is customary to put the integral number of divisions of the abscissa axis into correspondence with the number π, and then the fractions of the number π can be easily marked on the abscissa axis (it is especially convenient when the number of divisions of the number π is divisible by 12). Another scale, in which it is convenient to lay off tenths of a unity, is usually chosen for the axis of ordinates. The graphs obtained look deformed and the straight line $y = x$ is inclined to the abscissa axis at an angle of 45°.

Using millimetre graph paper and assuming 12 cm laid off along the abscissa axis to be π and 5 cm to be the scale unit laid off along the axis of ordinates, construct the graph of the functions $y = \sin x$ and $y = \cos x$.

62. In your squared copybook draw several periods of the graphs of the functions $y = \sin x$ and $y = \cos x$, assuming two squares of the sheet to be the scale unit laid off along the axis of ordinates and six squares laid off along the axis of abscissas to be π. Give a graphic interpretation of the solution of the following equations:

$$\sin x = 0.5, \quad \cos x = 0.5, \quad \sin x = \sqrt{2}/2, \quad \cos x = \sqrt{2}/2.$$

63. Using the same coordinate system, construct the schematic graphs of the following functions:

1. $y = \sin x$, $y = 2 \sin x$.
2. $y = \cos x$, $y = 3 \cos x$.
3. $y = \cos x$, $y = \frac{1}{3} \cos x$.
4. $y = \sin x$, $y = \frac{1}{2} \sin x$.

5. $y=\sin x,\ y=-\frac{1}{2}\sin x.$ 6. $y=\cos x,\ y=-2\cos x.$

7. $y=\cos x,\ y=\cos 2x.$ 8. $y=\sin x,\ y=\sin\frac{1}{2}x.$

9. $y=\sin x,\ y=\sin\left(\frac{1}{2}x-\frac{\pi}{6}\right).$

10. $y=\cos x,\ y=\cos\left(2x+\frac{2\pi}{3}\right).$

11. $y=\cos x,\ y=2\cos\left(\frac{x}{2}+\frac{\pi}{4}\right).$

12. $y=\sin x,\ y=\frac{1}{2}\sin\left(2x-\frac{\pi}{3}\right).$

64. Construct one period of the graphs of the following functions, first finding the period and choosing a convenient x-axis scale unit.

1. $y=\cos\pi x.$ 2. $y=\sin\pi x.$ 3. $y=2\sin\left(\frac{\pi}{2}x-\frac{\pi}{4}\right).$

4. $y=\frac{1}{2}\cos\left(\frac{\pi}{2}x+\frac{\pi}{4}\right).$ 5. $y=\frac{1}{2}\cos(2\pi x+\pi).$

6. $y=2\sin(2\pi x-\pi).$

65. Construct several periods of the graphs of the functions $y=\tan x$ and $y=\cot x$. First choose the same scale unit to be laid off along both axes, say, 1 cm, then construct the graphs to a more convenient scale, say, accepting 2 cm to be π and lay it off along the x-axis. Construct tangents to the graphs of the functions at the point $x=0$ for the function $\tan x$ and at the point $\pi/2$ for the function $\cot x$.

66. Using the constructed graphs of the functions $y=\tan x$ and $y=\cot x$, give a graphical interpretation of the following equations:

$$\tan x=1,\ \cot x=1,\ \tan x=-\sqrt{3},\ \cot x=-\sqrt{3}.$$

Chapter 5

TRANSFORMATION OF EXPRESSIONS WHICH INCLUDE TRIGONOMETRIC FUNCTIONS. SOME APPLICATIONS

5.1. The Sine and Cosine of the Sum and the Difference of Arguments

Trigonometric functions possess the following remarkable property: the so-called *addition formulas* exist for them which make it possible to express the values of the functions of the sum of the arguments in terms of the values of the functions of the terms. Namely, the following formulas hold true:

$$\sin(\varphi_1 + \varphi_2) = \sin\varphi_1 \cos\varphi_2 + \cos\varphi_1 \sin\varphi_2, \tag{1}$$

$$\cos(\varphi_1 + \varphi_2) = \cos\varphi_1 \cos\varphi_2 - \sin\varphi_1 \sin\varphi_2. \tag{2}$$

Let us prove the validity of these formulas. We first assume that both arguments φ_1 and φ_2 belong to the interval $(0, \pi/2)$, i.e. are measures of acute angles. We consider a circle with centre at the origin and radius 1 (Fig. 5.1). Let $\angle M_0OM_1 = \varphi_1$ and $\angle M_1OM_2 = \varphi_2$. We draw a perpendicular M_2K from the point M_2 to OM_1 and a perpendicular M_2N from the point M_2 to the abscissa axis. It is clear that $\angle NM_2K = \angle M_0OM_1 = \varphi_1$ since these angles are both acute and have mutually perpendicular sides. Then we draw perpendiculars KH and KL from the point K to the abscissa axis and to the segment M_2N. Considering the triangle OM_2K, we find that $OK = \cos\varphi_2$ and $M_2K = \sin\varphi_2$ since these segments are legs of a right triangle with an acute angle φ_2 and hypotenuse equal to 1. Furthermore, $LM_2 = M_2K\cos\varphi_1$, $LK = M_2K\sin\varphi_1$ since these segments are legs of a right triangle M_2KL with an acute angle φ_1 at the vertex M_2. Consequently, $LM_2 = \sin\varphi_2\cos\varphi_1$, $LK =$

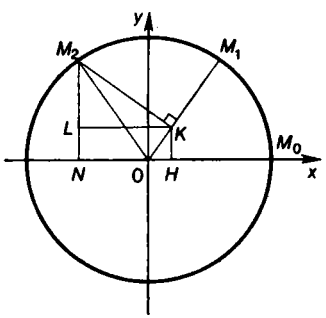


Fig. 5.1

$\sin \varphi_2 \sin \varphi_1$. Considering the triangle OKH, we find that

$$OH = OK \cos \varphi_1 = \cos \varphi_2 \cos \varphi_1,$$
$$KH = OK \sin \varphi_1 = \cos \varphi_2 \sin \varphi_1.$$

Finally,

$$\begin{aligned}\cos (\varphi_1 + \varphi_2) &= OH - NH = OH - LK \\ &= \cos \varphi_2 \cos \varphi_1 - \sin \varphi_2 \sin \varphi_1, \\ \sin (\varphi_1 + \varphi_2) &= M_2N = NL + LM_2 = HK + LM_2 \\ &= \cos \varphi_2 \sin \varphi_1 + \sin \varphi_2 \cos \varphi_1.\end{aligned}$$

We have thus proved formulas (1) and (2) for the acute angles φ_1 and φ_2.

It is evident that if formulas (1) and (2) hold true for some φ_1 and φ_2, they will remain true after the addition of the numbers $2k_1\pi$ and $2k_2\pi$ to φ_1 and φ_2 for any integers k_1 and k_2 since as a result of such an addition, φ_1, φ_2 and $\varphi_1 + \varphi_2$ change by the period of the sine and cosine functions. Therefore we can assume formulas (1) and (2) to have been proved for any φ_1 and φ_2 belonging to the first quadrant.

It is evident that any angle of the second quadrant can be obtained from that of the first quadrant by adding $\pi/2$, and the further addition of $\pi/2$ yields any angle of the third quadrant and, adding $\pi/2$ once more, we obtain any angle of the fourth quadrant. Therefore, proving the fact that if formulas (1) and (2) hold true for some φ_1 and φ_2, then they obviously remain true after the addition of $\pi/2$ to one of the terms, then we shall prove formulas (1) and (2) for any φ_1 and φ_2 since any φ_1 and φ_2 can be obtained from the angles of the first quadrant by means of the addition of $\pi/2$ several times to each of the arguments.

Thus it remains to prove the following lemma:

Lemma. *If formulas* (1) *and* (2) *hold true for some* φ_1 *and* φ_2, *then they remain true after the addition of* $\frac{\pi}{2}$ *to one of the arguments.*

Proof. Assume that (1) and (2) hold true for φ_1 and φ_2 and set

$$\varphi_1' = \frac{\pi}{2} + \varphi_1.$$

We have

$$\begin{aligned}\sin(\varphi_1' + \varphi_2) &= \sin\left(\frac{\pi}{2} + \varphi_1 + \varphi_2\right) = \cos(\varphi_1 + \varphi_2)\\ &= \cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2\\ &= \cos\left(\varphi_1' - \frac{\pi}{2}\right)\cos\varphi_2 - \sin\left(\varphi_1' - \frac{\pi}{2}\right)\sin\varphi_2\\ &= \sin\varphi_1'\cos\varphi_2 + \cos\varphi_1'\sin\varphi_2.\end{aligned}$$

Similarly,

$$\begin{aligned}\cos(\varphi_1' + \varphi_2) &= \cos\left(\frac{\pi}{2} + \varphi_1 + \varphi_2\right) = -\sin(\varphi_1 + \varphi_2)\\ &= -\sin\varphi_1\cos\varphi_2 - \cos\varphi_1\sin\varphi_2\\ &= -\sin\left(\varphi_1' - \frac{\pi}{2}\right)\cos\varphi_2 - \cos\left(\varphi_1' - \frac{\pi}{2}\right)\sin\varphi_2\\ &= \cos\varphi_1'\cos\varphi_2 - \sin\varphi_1'\sin\varphi_2.\end{aligned}$$

We have proved the lemma.

Formulas (1) and (2) immediately yield formulas

$$\sin(\varphi_1 - \varphi_2) = \sin\varphi_1\cos\varphi_2 - \cos\varphi_1\sin\varphi_2, \tag{3}$$

$$\cos(\varphi_1 - \varphi_2) = \cos\varphi_1\cos\varphi_2 + \sin\varphi_1\sin\varphi_2. \tag{4}$$

They result from (1) and (2) after the substitution of $-\varphi_2$ for φ_2.

Here is another proof of formulas (1)-(4) based on the theory of vectors. Recall that the *scalar product* of two vectors is the product of the lengths of those vectors by the cosine of the angle between them. In terms of the coordinates of vectors (i.e. in terms of the projections of the vectors on the coordinate axes) a scalar product can be expressed according to the formula $x_1x_2 + y_1y_2$, where x_1, y_1 and x_2, y_2 are the coordinates of these vectors. Assume that r_1 and r_2 are the lengths of the given vectors and φ_1 and φ_2 are the angles the vectors make with the abscissa axis (the angles being understood in the generalized sense). Then $x_1 = r_1\cos\varphi_1$, $y_1 = r_1\sin\varphi_1$, $x_2 = r_2\cos\varphi_2$ and $y_2 = r_2\sin\varphi_2$. The angle between the vectors is equal to $\varphi_1 - \varphi_2$ or $\varphi_2 - \varphi_1$ with an accuracy to within the period so that the cosine of the angle between the vectors is equal to $\cos(\varphi_1 - \varphi_2)$. Comparing the two forms of the notation of a scalar product (in terms of the cosine of the angle between the vectors and in terms of

the coordinates of those vectors), we arrive at a relation

$$r_1 r_2 \cos(\varphi_1 - \varphi_2) = r_1 \cos\varphi_1 \cdot r_2 \cos\varphi_2 + r_1 \sin\varphi_1 r_2 \sin\varphi_2,$$

whence, after the division by $r_1 r_2$, we get formula (4):

$$\cos(\varphi_1 - \varphi_2) = \cos\varphi_1 \cos\varphi_2 + \sin\varphi_1 \sin\varphi_2.$$

It was derived without any restrictions imposed on φ_1 and φ_2.

Formula (4) easily yields formulas (2), (1) and (3). The substitution of $-\varphi_2$ for φ_2 in formula (4) yields (2):

$$\cos(\varphi_1 + \varphi_2) = \cos\varphi_1 \cos\varphi_2 - \sin\varphi_1 \sin\varphi_2.$$

Replacing the argument φ_1 by $\frac{\pi}{2} - \varphi_1$ in (4), we obtain

$$\cos\left(\frac{\pi}{2} - \varphi_1 - \varphi_2\right) = \cos\left(\frac{\pi}{2} - \varphi_1\right)\cos\varphi_2 + \sin\left(\frac{\pi}{2} - \varphi_1\right)\sin\varphi_2,$$

whence follows (1):

$$\sin(\varphi_1 + \varphi_2) = \sin\varphi_1 \cos\varphi_2 + \cos\varphi_1 \sin\varphi_2.$$

Finally, replacing the argument φ_2 by $-\varphi_2$ in (1), we get (3):

$$\sin(\varphi_1 - \varphi_2) = \sin\varphi_1 \cos\varphi_2 - \cos\varphi_1 \sin\varphi_2.$$

Example. Find $\sin 15°$ and $\cos 15°$.

Solution.

$$\sin 15° = \sin(45° - 30°) = \sin 45° \cos 30° - \cos 45° \sin 30° = \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{3}}{2} - \frac{\sqrt{2}}{2 \cdot 2} = \frac{\sqrt{6} - \sqrt{2}}{4}.$$

$$\cos 15° = \cos(45° - 30°) = \cos 45° \cos 30° + \sin 45° \sin 30° = \frac{\sqrt{2}}{2} \cdot \frac{\sqrt{3}}{2} + \frac{\sqrt{2} \cdot 1}{2 \cdot 2} = \frac{\sqrt{6} + \sqrt{2}}{4}.$$

5.2. The Tangent and Cotangent of a Sum and a Difference

The following formulas hold true for the tangent and the cotangent:

$$\tan(\varphi_1 + \varphi_2) = \frac{\tan\varphi_1 + \tan\varphi_2}{1 - \tan\varphi_1 \tan\varphi_2}, \tag{5}$$

$$\cot(\varphi_1 + \varphi_2) = \frac{\cot\varphi_1 \cot\varphi_2 - 1}{\cot\varphi_1 + \cot\varphi_2}. \tag{6}$$

Indeed,

$$\tan(\varphi_1+\varphi_2)=\frac{\sin(\varphi_1+\varphi_2)}{\cos(\varphi_1+\varphi_2)}$$
$$=\frac{\sin\varphi_1\cos\varphi_2+\cos\varphi_1\sin\varphi_2}{\cos\varphi_1\cos\varphi_2-\sin\varphi_1\sin\varphi_2}.$$

Dividing both the numerator and the denominator by $\cos\varphi_1\cos\varphi_2$, we obtain

$$\tan(\varphi_1+\varphi_2)=\frac{\frac{\sin\varphi_1}{\cos\varphi_1}+\frac{\sin\varphi_2}{\cos\varphi_2}}{1-\frac{\sin\varphi_1\sin\varphi_2}{\cos\varphi_1\cos\varphi_2}}=\frac{\tan\varphi_1+\tan\varphi_2}{1-\tan\varphi_1\tan\varphi_2}.$$

Note that the right-hand side of this relation is meaningless for $\varphi_1=\pi/2$ or $\varphi_2=\pi/2$ (with an accuracy of an integer which is a multiple of π) whereas the left-hand side has sense. Prior to the division by $\cos\varphi_1\cdot\cos\varphi_2$ the right-hand side had sense for these values of φ_1 and φ_2 as well, they lose sense only as a result of the division by $\cos\varphi_1\cdot\cos\varphi_2$, i.e. by the expression that vanishes for $\varphi_1=\frac{\pi}{2}+\pi k$ or $\varphi_2=\frac{\pi}{2}+\pi k$, where k is any integer.

In the same way we can derive a formula for the cotangent of a sum:

$$\cot(\varphi_1+\varphi_2)=\frac{\cos(\varphi_1+\varphi_2)}{\sin(\varphi_1+\varphi_2)}$$
$$=\frac{\cos\varphi_1\cos\varphi_2+\sin\varphi_1\sin\varphi_2}{\sin\varphi_1\cos\varphi_2+\cos\varphi_1\sin\varphi_2}.$$

After the division of the numerator and denominator by $\sin\varphi_1\cdot\sin\varphi_2$, we obtain

$$\cot(\varphi_1+\varphi_2)=\frac{\cot\varphi_1\cot\varphi_2-1}{\cot\varphi_2+\cot\varphi_1}.$$

The substitution of $-\varphi_2$ for φ_2 leads to formulas

$$\tan(\varphi_1-\varphi_2)=\frac{\tan\varphi_1-\tan\varphi_2}{1+\tan\varphi_1\tan\varphi_2}, \tag{7}$$

$$\cot(\varphi_1-\varphi_2)=\frac{\cot\varphi_1\cot\varphi_2+1}{\cot\varphi_2-\cot\varphi_1}. \tag{8}$$

Example 1. Let $\tan\varphi_1=1/2$ and $\tan\varphi_2=1/3$. What is $\tan(\varphi_1+\varphi_2)$ equal to?

Solution.

$$\tan(\varphi_1+\varphi_2)=\frac{\frac{1}{2}+\frac{1}{3}}{1-\frac{1}{2}\cdot\frac{1}{3}}=1.$$

Thus, if φ_1 and φ_2 are acute angles, then $\varphi_1+\varphi_2=\pi/4$.

5.3. Trigonometric Functions of a Doubled Argument and of Some Multiple Arguments

Setting $\varphi_1=\varphi_2=\varphi$ in the formulas for the addition of arguments, we obtain

$$\sin 2\varphi=\sin\varphi\cos\varphi+\cos\varphi\sin\varphi=2\sin\varphi\cos\varphi,$$

$$\cos 2\varphi=\cos\varphi\cos\varphi-\sin\varphi\sin\varphi=\cos^2\varphi-\sin^2\varphi,$$

$$\tan 2\varphi=\frac{\tan\varphi+\tan\varphi}{1-\tan\varphi\tan\varphi}=\frac{2\tan\varphi}{1-\tan^2\varphi},$$

$$\cot 2\varphi=\frac{\cot\varphi\cot\varphi-1}{\cot\varphi+\cot\varphi}=\frac{\cot^2\varphi-1}{2\cot\varphi}.$$

Example 1. Express $\cos 3\varphi$ in terms of $\cos\varphi$.

Solution.

$$\begin{aligned}\cos 3\varphi &= \cos(2\varphi+\varphi)=\cos 2\varphi\cos\varphi-\sin 2\varphi\sin\varphi\\ &=(\cos^2\varphi-\sin^2\varphi)\cos\varphi-2\sin\varphi\cos\varphi\sin\varphi\\ &=\cos^3\varphi-3\sin^2\varphi\cos\varphi\\ &=\cos^3\varphi-3(1-\cos^2\varphi)\cos\varphi\\ &=4\cos^3\varphi-3\cos\varphi.\end{aligned}$$

◀ **Example 2.** Use the result of Example 1 to solve the equation $x^3-3x-a=0$ for $|a|<2$ with the aid of trigonometric functions.

Solution. We set $x=2\cos\varphi$. Then the equation assumes the form $8\cos^3\varphi-6\cos\varphi-a=0$, i.e. $4\cos^3\varphi-3\cos\varphi=a/2$ and, finally, $\cos 3\varphi=a/2$. By the hypothesis $|a|<2$, and so $|a/2|<1$. Consequently, there is an angle α_0 whose cosine is equal to $a/2$. It is defined with an accuracy of the sign and of the integral multiple of $2\pi=360°$. It follows from the equation $\cos 3\varphi=a/2$ that $3\varphi=\pm(\alpha_0+k\cdot 360°)$ and $\varphi=\pm\left(\frac{\alpha_0}{3}+k\cdot 120°\right)$, where k is any integer. The roots of the initial equation can be found

from the formula

$$x = 2\cos \pm \left(\frac{\alpha_0}{3} + k \cdot 120^\circ\right) = 2\cos\left(\frac{\alpha_0}{3} + k \cdot 120^\circ\right).$$

It is sufficient to take the values 0, 1, 2 for the integer k since in the case of $k = 3$ we shall have the same value for x as in the case of $k = 0$, in the case of $k = 4$, the same value as in the case of $k = 1$, in the case of $k = 5$, the same value as in the case of $k = 2$ and so on. We have thus found three roots of the equation. These roots are pairwise distinct since if

$$\cos\left(\frac{\alpha_0}{3} + k_1 \cdot 120^\circ\right) = \cos\left(\frac{\alpha_0}{3} + k_2 \cdot 120^\circ\right)$$

for $k_1 \neq k_2$ and $0 \leqslant k_1 \leqslant 2$, $0 \leqslant k_2 \leqslant 2$, then either

$$\frac{\alpha_0}{3} + k_1 \cdot 120^\circ = \frac{\alpha_0}{3} + k_2 \cdot 120^\circ,$$

what is impossible, or

$$\frac{\alpha_0}{3} + k_1 \cdot 120^\circ = -\left(\frac{\alpha_0}{3} + k_2 \cdot 120^\circ\right) + k_3 \cdot 360^\circ,$$

whence $2\alpha_0$ is an integer which is a multiple of 360° and, with an accuracy to within a period, α_0 is 0° or 180°, and so $a = 2\cos\alpha = \pm 2$, and this contradicts the assumption that $|a| < 2$. For $a = \pm 2$ there is indeed a multiple root since

$$x^3 - 3x - 2 = (x + 1)^2 (x - 2)$$

and

$$x^3 - 3x + 2 = (x - 1)^2 (x + 2).$$

Since the initial equation is a cubic one and cannot have more than three roots, we have found all the roots.

Note that for $|a| < 2$, we can reduce a more general equation $x^3 - px - q = 0$, where $p > 0$ and $\frac{q^2}{4} < \frac{p^3}{27}$, to the form $x^3 - 3x - a = 0$. To do this, we must set $x = \sqrt{\frac{p}{3}} \cdot y$ and divide by $(\sqrt{p/3})^3$.

Example 3. Prove that $\cos kx$ and $\frac{\sin kx}{\sin x}$ can be expressed as a polynomial of $\cos x$ for any natural k.

Solution. We apply induction. For $k = 1$ the assertion is obvious. Assume that we have established that

$$\cos kx = P_k(\cos x), \quad \sin kx = \sin x Q_{k-1}(\cos x),$$

where P_k and Q_{k-1} are some polynomials. We shall show that similar formulas hold true for $\cos(k+1)x$ and $\sin(k+1)x$. We have

$$\begin{aligned}\cos(k+1)x &= \cos(kx+x) = \cos kx\cos x - \sin kx\sin x\\ &= P_k(\cos x)\cos x - \sin x Q_{k-1}(\cos x)\sin x\\ &= P_k(\cos x)\cos x - (1-\cos^2 x)\,Q_{k-1}(\cos x)\\ &= P_{k+1}(\cos x),\end{aligned}$$

where $P_{k+1}(z) = P_k(z)\cdot z + (z^2-1)\,Q_{k-1}(z)$. Furthermore,

$$\begin{aligned}\sin(k+1)x &= \sin(kx+x) = \sin kx\cos x + \cos(kx)\sin x\\ &= \sin x Q_{k-1}(\cos x)\cos x + P_k(\cos x)\sin x\\ &= \sin x\,(Q_{k-1}(\cos x)\cos x + P_k(\cos x))\\ &= \sin x Q_k(\cos x),\end{aligned}$$

where $Q_k(z) = Q_{k-1}(z)\cdot z - P_k(z)$. It immediately follows from these formulas that k and $k-1$ are equal to the powers of the polynomials P_k and Q_{k-1} respectively.

Our arguments not only prove the existence of the polynomials $P_k(\cos x)$ and $Q_{k-1}(\cos x)$ but also make it possible to calculate them successively. Evidently, $P_1(z) = z$ and $Q_0(z) = 1$. Furthermore,

$$\begin{aligned}P_2(z) &= P_1(z)\,z + (z^2-1)\,Q_0(z) = 2z^2-1,\\ Q_1(z) &= Q_0(z)\,z + P_1(z) = 2z;\\ P_3(z) &= P_2(z)\,z + (z^2-1)\,Q_1(z) = 4z^3-3z,\\ Q_2(z) &= Q_1(z)\,z + P_2(z) = 4z^2-1.\end{aligned}$$

The successive calculations yield

$$P_4(z) = 8z^4 - 8z^2 + 1,\quad Q_3(z) = 8z^3 - 4z,$$
$$P_5(z) = 16z^5 - 20z^3 + 5z,\quad Q_4(z) = 16z^4 - 12z^2 + 1$$

and so on.

Thus, for instance,

$$\cos 5x = 16\cos^5 x - 20\cos^3 x + 5\cos x,$$
$$\sin 5x = \sin x\,(16\cos^4 x - 12\cos^2 x + 1). \blacktriangleright$$

5.4. Trigonometric Functions of a Half-Argument

By means of addition and subtraction we infer from the formulas $\cos^2 \varphi - \sin^2 \varphi = \cos 2\varphi$ and $\cos^2 \varphi + \sin^2 \varphi = 1$ that

$$\cos^2 \varphi = \frac{1+\cos 2\varphi}{2}, \quad \sin^2 \varphi = \frac{1-\cos 2\varphi}{2}.$$

Setting $\varphi = x/2$, we obtain

$$\cos^2 \frac{x}{2} = \frac{1+\cos x}{2}, \quad \sin^2 \frac{x}{2} = \frac{1-\cos x}{2},$$

whence it follows that

$$\cos \frac{x}{2} = \pm \sqrt{\frac{1+\cos x}{2}}, \quad \sin \frac{x}{2} = \pm \sqrt{\frac{1-\cos x}{2}}.$$

The signs are chosen in accordance with the quadrant to which $x/2$ belongs.

To calculate $\tan (x/2)$, we apply a more artificial technique, namely,

$$\tan \frac{x}{2} = \frac{\sin (x/2)}{\cos (x/2)} = \frac{2 \sin (x/2) \cos (x/2)}{2 \cos^2 (x/2)} = \frac{\sin x}{1+\cos x}.$$

Similarly, multiplying the numerator and denominator of the fraction $\frac{\sin (x/2)}{\cos (x/2)}$ by $2 \sin (x/2)$, we get $\tan \frac{x}{2} = \frac{2 \sin^2 (x/2)}{2 \sin (x/2) \cos (x/2)} = \frac{1-\cos x}{\sin x}$.

Example 1. Find $\tan \frac{\pi}{8}$.

Solution.

$$\tan \frac{\pi}{8} = \tan \frac{1}{2} \cdot \frac{\pi}{4} = \frac{1-\cos (\pi/4)}{\sin (\pi/4)} = \frac{1-\sqrt{2}/2}{\sqrt{2}/2} = \sqrt{2}-1.$$

Example 2. Find $\sin (\pi/8)$ and $\cos (\pi/8)$.

Solution.

$$\sin^2 \frac{\pi}{2} = \frac{1-\cos \frac{\pi}{4}}{2} = \frac{1-\frac{\sqrt{2}}{2}}{2} = \frac{2-\sqrt{2}}{4},$$

$$\sin \frac{\pi}{8} = \sqrt{\frac{2-\sqrt{2}}{2}},$$

$$\cos^2 \frac{\pi}{8} = \frac{1+\cos \frac{\pi}{4}}{2} = \frac{1+\frac{\sqrt{2}}{2}}{2} = \frac{2+\sqrt{2}}{4},$$

$$\cos \frac{\pi}{8} = \sqrt{\frac{2+\sqrt{2}}{2}}.$$

In both cases, when we extract the square root, we take the plus sign since $\pi/8$ is an angle belonging to the first quadrant.

◀ 5.5. Expressing Trigonometric Functions in Terms of the Tangent of a Half-Argument

We shall show that $\sin x$, $\cos x$, $\tan x$ and $\cot x$ can be expressed as rational functions of $\tan (x/2)$. Consequently, any rational expression of $\sin x$, $\cos x$, $\tan x$ and $\cot x$ (and also of multiple arguments) can be expressed as rational functions of $\tan (x/2)$.

We have $\sin x = 2 \sin (x/2) \cos (x/2) = 2 \tan (x/2) \cos^2 (x/2)$. But $\cos^2 (x/2) = \dfrac{1}{1+\tan^2 (x/2)}$. Thus $\sin x = \dfrac{2 \tan (x/2)}{1+\tan^2 (x/2)}$. Furthermore,

$$\tan x = \frac{2 \tan (x/2)}{1-\tan^2 (x/2)}, \quad \cos x = \frac{\sin x}{\tan x} = \frac{1-\tan^2 (x/2)}{1+\tan^2 (x/2)}.$$

And finally,

$$\cot x = \frac{1}{\tan x} = \frac{1-\tan^2 (x/2)}{2 \tan (x/2)}.$$

Note that the right-hand sides of these formulas lose sense for $x = \pi + 2k\pi$, where k is an integer.

Example 1. Express $3 \sin x + 4 \cos x$ in terms of $\tan (x/2)$.

Solution.

$$3 \sin x + 4 \cos x = \frac{6 \tan (x/2) + 4 - 4 \tan^2 (x/2)}{1+\tan^2 (x/2)}$$

$$= \frac{2 (2-\tan (x/2)) (1+2 \tan (x/2))}{2+\tan^2 (x/2)}. \blacktriangleright$$

5.6. Expressing the Products of the Sine and Cosine Functions as Sums and Expressing Sums as Products

We have two groups of formulas:

$$\left.\begin{aligned}
\sin x \cos y &= \frac{\sin (x+y)+\sin (x-y)}{2}, \\
\cos x \cos y &= \frac{\cos (x+y)+\cos (x-y)}{2}, \\
\sin x \sin y &= \frac{\cos (x-y)-\cos (x+y)}{2},
\end{aligned}\right\} \quad \text{(I)}$$

$$\left.\begin{aligned}
\sin \alpha+\sin \beta &= 2 \sin \frac{\alpha+\beta}{2} \cos \frac{\alpha-\beta}{2}, \\
\sin \alpha-\sin \beta &= 2 \sin \frac{\alpha-\beta}{2} \cos \frac{\alpha+\beta}{2}, \\
\cos \alpha+\cos \beta &= 2 \cos \frac{\alpha+\beta}{2} \cos \frac{\alpha-\beta}{2}, \\
\cos \alpha-\cos \beta &= -2 \sin \frac{\alpha+\beta}{2} \sin \frac{\alpha-\beta}{2}.
\end{aligned}\right\} \quad \text{(II)}$$

It is very easy to derive these formulas. Adding and subtracting the formulas for the sine of the sum and the difference and for the cosine of the sum and the difference, we arrive at formulas

$$\left.\begin{aligned}
\sin (x+y)+\sin (x-y) &= 2 \sin x \cos y, \\
\sin (x+y)-\sin (x-y) &= 2 \cos x \sin y, \\
\cos (x-y)+\cos (x+y) &= 2 \cos x \cos y, \\
\cos (x+y)-\cos (x-y) &= -2 \sin x \sin y.
\end{aligned}\right\} \quad \text{(III)}$$

Formulas of group I can be obtained from the first, the third and the fourth formula of group III as a result of the division of both sides by 2.

To derive formulas of group II, we must set $x+y=\alpha$, $x-y=\beta$, whence $x=\frac{\alpha+\beta}{2}$ and $y=\frac{\alpha-\beta}{2}$.

Example 1. Transform $\sin x \sin 2x \sin 3x$ into a sum.

Solution.

$$\sin x \sin 2x = \frac{\cos(-x) - \cos 3x}{2} = \frac{1}{2}(\cos x - \cos 3x),$$

$$\sin x \sin 2x \sin 3x = \frac{1}{2}(\sin 3x \cos x - \sin 3x \cos 3x)$$

$$= \frac{1}{4}(\sin 4x + \sin 2x - \sin 6x).$$

Example 2. Transform

$$\sin \frac{x}{2}(\sin x + \sin 2x + \sin 3x + \sin 4x + \sin 5x).$$

Solution.

$$\sin \frac{x}{2} \sin x = \frac{1}{2}\left(\cos \frac{x}{2} - \cos \frac{3}{2} x\right),$$

$$\sin \frac{x}{2} \sin 2x = \frac{1}{2}\left(\cos \frac{3}{2} x - \cos \frac{5}{2} x\right),$$

$$\sin \frac{x}{2} \sin 3x = \frac{1}{2}\left(\cos \frac{5}{2} x - \cos \frac{7}{2} x\right),$$

$$\sin \frac{x}{2} \sin 4x = \frac{1}{2}\left(\cos \frac{7}{2} x - \cos \frac{9}{2} x\right),$$

$$\sin \frac{x}{2} \sin 5x = \frac{1}{2}\left(\cos \frac{9}{2} x - \cos \frac{11}{2} x\right),$$

whence it follows that

$$\sin \frac{x}{2}(\sin x + \sin 2x + \sin 3x + \sin 4x + \sin 5x)$$

$$= \frac{1}{2}\left(\cos \frac{1}{2} x - \cos \frac{11}{2} x\right) = \frac{1}{2} \cdot 2 \sin 3x \sin \frac{5}{2} x$$

$$= \sin 3x \sin \frac{5}{2} x.$$

Thus we have found that

$$\sin x + \sin 2x + \sin 3x + \sin 4x + \sin 5x$$

$$= \frac{1}{\sin(x/2)} \sin 3x \sin \frac{5}{2} x.$$

In a like manner we can "contract" the sum of the values of the sines and cosines whose arguments form an arithmetic progression.

5.7. Transformation of a Linear Combination of a Sine and a Cosine

In applications of mathematics we often encounter a function of the form $a \sin x + b \cos x$ with nonzero coefficients a and b. It is a periodic function with period 2π. The absolute values of the terms are sometimes summed up when the terms are of like sign and sometimes are subtracted when the signs are unlike. To comprehend the variation of the function, we must transform it. We first assume that $a = \cos \alpha$ and $b = \sin \alpha$. Then

$$a \sin x + b \cos x = \sin x \cos \alpha + \cos x \sin \alpha$$

$$= \sin (x + \alpha).$$

After this transformation it becomes clear that the graph of the function is a sine curve displaced by $-\alpha$ along the abscissa axis.

It stands to reason that we cannot always find an α such that $a = \cos \alpha$, $b = \sin \alpha$. To find such an α, it is necessary and sufficient that $a^2 + b^2 = 1$. Indeed, this condition means that the point M (a, b) lies on a circle of radius 1 with centre at the origin and its coordinates a and b are the cosine and sine of the angle the vector OM makes with the positive direction of the abscissa axis.

However, we can always find r and α such that $a = r \cos \alpha$, $b = r \sin \alpha$. To do this, it is sufficient to take $r = \sqrt{a^2 + b^2}$ since with this choice of r the numbers $a_1 = a/r$ and $b_1 = b/r$ satisfy the requirement $a_1^2 + b_1^2 = 1$ and, consequently, there is an α such that $a_1 = a/r = \cos \alpha$, $b_1 = b/r = \sin \alpha$. Substituting the numbers $r \cos \alpha$ and $r \sin \alpha$ for a and b, we get

$$a \sin x + b \cos x = r \sin x \cos \alpha + r \cos x \sin \alpha$$

$$= r \sin (x + \alpha).$$

It is clear from this representation that the graph of the function $a \cos x + b \sin x$ is a sine curve with an amplitude $r = \sqrt{a^2 + b^2}$ displaced along the abscissa axis by $-\alpha$.

Example 1.

$$\sin x - \cos x = \sqrt{2}\left(\frac{1}{\sqrt{2}}\sin x - \frac{1}{\sqrt{2}}\cos x\right)$$
$$= \sqrt{2}\left(\cos\left(-\frac{\pi}{4}\right)\sin x + \sin\left(-\frac{\pi}{4}\right)\cos x\right)$$
$$= \sqrt{2}\sin\left(x - \frac{\pi}{4}\right).$$

5.8. Harmonic Oscillations

The motion of a point described by the equation $y = b \sin(at + \alpha)$, $a > 0$, $0 \leqslant \alpha < 2\pi$, is known as *harmonic oscillation.* In general, we can say about any quantity that varies with time according to the law $y = b \sin(at + \alpha)$, that it oscillates harmonically. A quantity which oscillates harmonically deviates periodically from zero to both sides by a quantity $|b|$, which is called an *amplitude of the oscillation.* The smallest period is evidently $\frac{2\pi}{a}$. The inverse quantity $\frac{a}{2\pi}$ is the *frequency of oscillation.* This number shows how many complete periods of oscillations the point makes in unit time.

In physics and engineering frequency is measured in hertz. One hertz is the frequency of oscillation with one period per second. In radioengineering we have to deal with oscillations of a very high frequency and then kilohertz is a more convenient unit (1 kHz $= 10^3$ Hz) and even a megahertz (1 MHz $= 10^6$ Hz). The number α is the *initial phase of a harmonic oscillation.* Without losing generality, we can assume that $b > 0$ in the equation $y = b \sin(at + \alpha)$ since a change of sign of b is equivalent to the addition of π or $-\pi$ to the initial phase.

A harmonic oscillation is the simplest but convenient mathematical model to describe oscillation processes occurring in nature and engineering. A load attached to an ideal spring moves according to the law of harmonic oscillation when it is disturbed from the equilibrium condition (if the air drag is neglected). This law can also describe the oscillations of an electromotive force in an a.c. generator, electromagnetic oscillations in radioengineering. This law is also used in the description of the nature of light, of the

oscillation of every point of a stretched string disturbed from the equilibrium condition, of the oscillations of the molecules of a medium in which a sound propagates, etc.

Addition of harmonic oscillations is of interest. It occurs, in particular, when a.c. generators are connected to a circuit, when a sound comes from two closely spaced sources, etc. The sum of two harmonic oscillations $b_1 \sin (a_1 t + \alpha_1) + b_2 \sin (a_2 t + \alpha_2)$ describes, in general, very complicated oscillations, not necessarily periodic, in the case when the frequencies are noncommensurable. However, the following theorem holds true.

Theorem. *If the frequencies of two harmonic oscillations coincide, then the sum of these oscillations is again a harmonic oscillation of the same frequency.*

Proof. Let $y_1 = b_1 \sin (at + \alpha_1)$, $y_2 = b_2 \sin (at + \alpha_2)$, $a > 0$. We set $b_1 \geqslant b_2 > 0$. Then

$$\begin{aligned} y_1 + y_2 &= b_1 \sin (at + \alpha_1) + b_2 \sin (at + \alpha_2) \\ &= b_1 \sin at \cos \alpha_1 + b_1 \cos at \sin \alpha_1 \\ &\quad + b_2 \sin at \cos \alpha_2 + b_2 \cos at \sin \alpha_2 \\ &= (b_1 \cos \alpha_1 + b_2 \cos \alpha_2) \sin at \\ &\quad + (b_1 \sin \alpha_1 + b_2 \sin \alpha_2) \cos at. \end{aligned}$$

In accordance with the calculations carried out in the preceding section, we have $y_1 + y_2 = r \sin (at + \alpha_3)$,

$$\begin{aligned} r^2 &= (b_1 \cos \alpha_1 + b_2 \cos \alpha_2)^2 + (b_1 \sin \alpha_1 + b_2 \sin \alpha_2)^2 \\ &= b_1^2 \cos^2 \alpha_1 + 2b_1 b_2 \cos \alpha_1 \cos \alpha_2 + b_2^2 \cos^2 \alpha_2 \\ &\quad + b_1^2 \sin^2 \alpha_1 + 2b_1 b_2 \sin \alpha_1 \sin \alpha_2 + b_2^2 \sin^2 \alpha_2 \\ &= b_1^2 + 2b_1 b_2 \cos (\alpha_1 - \alpha_2) + b_2^2, \end{aligned}$$

and α_3 is defined by the relations

$$\cos \alpha_3 = \frac{b_1 \cos \alpha_1 + b_2 \cos \alpha_2}{r}, \quad \sin \alpha_3 = \frac{b_1 \sin \alpha_1 + b_2 \sin \alpha_2}{r}.$$

We have proved the theorem.

Note that the amplitude of the oscillation $y_1 + y_2$ depends not only on the amplitudes b_1 and b_2 of the summands of oscillation but also on the cosine of the displacement $\alpha_1 - \alpha_2$ of the initial phase. The amplitude is at a maximum, equal to $b_1 + b_2$, for $\alpha_1 = \alpha_2$ and at a minimum, equal to $b_1 - b_2$, for $\alpha_1 - \alpha_2 = \pi$.

5.9. Oscillations with a Variable Amplitude

Assume that $f(t)$ is a function of time t varying slowly as compared with the frequency $\frac{a}{2\pi}$ of the oscillation defined by the equation $y = \sin(at + \alpha)$ so that during one period of variation of y the values of the function $f(t)$ vary but little. Then the coordinate of a moving point with an equation of motion $z = f(t) \sin(at + \alpha)$ will vary from $f(t_1)$ to $-f(t_2)$, from $-f(t_2)$ to $f(t_3)$, from $f(t_3)$ to $-f(t_4)$ and so on, where $t_1, t_3, t_5, \ldots$ are the values of t for which $at + \alpha = \frac{\pi}{2} + 2k\pi$, and $t_2, t_4, t_6, \ldots$ are the values of t for which $at + \alpha = \frac{3\pi}{2} + 2k\pi$.

It is easy to see that the points $t_1, t_3, t_5, \ldots$ are not, in general, points of maximum for the function $z = f(t)\sin(at+\alpha)$ and, correspondingly, the points $t_2, t_4, t_6, \ldots$ are not points of minimum. To make sure of this, we shall calculate the value of the derivative of z at some point, say, the point t_1. We can do this although we do not yet know how to find a derivative of the sine. Indeed, t_1 is a point of maximum for $\sin(at + \alpha)$ and the derivative of $\sin(at+\alpha)$ is equal to 0. Thus $z'_{t=t_1} = f'(t_1) \sin(at_1 + \alpha) + f(t_1) \times [\sin(at + \alpha)]'_{t=t_1} = f'(t_1)$. It follows from this relation that the graph of the function $z = f(t) \sin(at + \alpha)$ has a tangent in common with the graph of the function $f(t)$. The same is true of the points $t_3, t_5, \ldots$, and at the points $t_2, t_4, \ldots$, z' differs only by sign from the corresponding values of f'. Thus, if, for instance, $f(t)$ decreases, then in the vicinity of the points $t_1, t_3, t_5, \ldots$ the graph of $z = f(t) \sin(at + \alpha)$ has the shape shown in Fig. 5.2 and the maxima of the function z are slightly to the left of the points $t_1, t_3, t_5, \ldots$. At the points of maximum the values of

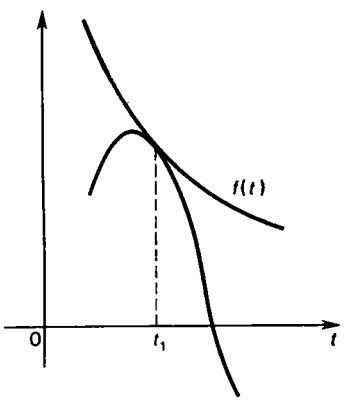


Fig. 5.2

z are a little smaller than those of $f(t)$ at these points. Thus the motion defined by the equation $z = f(t) \sin(at + \alpha)$ can be regarded as an oscillatory motion with a variable amplitude. In particular, damped oscillations with an equation $z = e^{-kt} \sin(at + \alpha)$, $k > 0$, are of this type. It is this law that governs the motion of a load attached to a spring, when the air drag is taken into account, or a point of a string oscillating in the air which hampers its motion.

An interesting and important case of oscillations with a variable amplitude occurs when harmonic oscillations

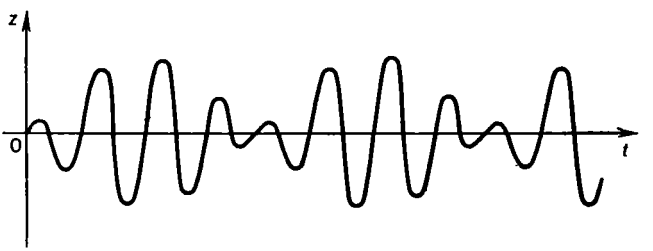


Fig. 5.3

with the same amplitude and different but very close frequencies are added together.

Assume that $y_1 = b \sin(a_1 t + \alpha_1)$, $y_2 = b \sin(a_2 t + \alpha_2)$ and $a_1 - a_2$ is a positive number much smaller than a_1 and a_2. Then

$$z = y_1 + y_2 = b(\sin(a_1 t + \alpha_1) + \sin(a_2 t + \alpha_2))$$
$$= b \sin\left(\frac{a_1 + a_2}{2} t + \frac{\alpha_1 + \alpha_2}{2}\right) \cos\left(\frac{a_1 - a_2}{2} t + \frac{\alpha_1 - \alpha_2}{2}\right).$$

The first factor oscillates with the frequency $\frac{1}{2\pi} \cdot \frac{a_1 + a_2}{2}$ which is close to the frequencies of the oscillations y_1 and y_2. The second factor varies much slower, oscillating with a much lower frequency $\frac{1}{2\pi} \cdot \frac{a_1 - a_2}{2}$.

Therefore $z = y_1 + y_2$ specifies an oscillatory motion with frequency close to that of y_1 and y_2 but with a variable amplitude which makes oscillatory motions of an appreciably lower frequency. The graph of this motion resembles that shown in Fig. 5.3. A phenomenon occurs which is known in physics as *beats*.

5.10. Basic Trigonometric Equations and Inverse Trigonometric Functions

Let us discuss the equation $\sin \varphi = y$ with respect to φ. The solution of this equation is $|y| \leqslant 1$. When we represent the cosine and sine functions as the coordinates of a point on the circumference of a circle, corresponding to the angle φ, we see (Fig. 5.4) that for $y \neq \pm 1$ there are two points, M and M', with the given ordinate y on the circle, M on the right of the y-axis and M' on the left of it.

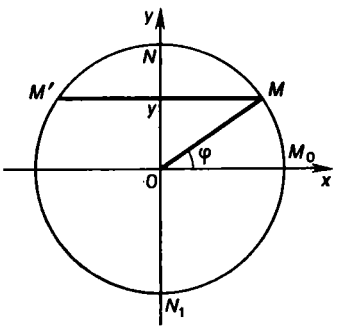


Fig. 5.4

We shall first consider the point M which is on the right of the y-axis. As y increases from 0 to 1, the point M continuously moves along the circle from the point M_0 to the point N. By means of a simple reckoning, find that the corresponding angle continuously increases from 0 to $\pi/2$. As y varies from 0 to -1, the point M continuously moves in the negative direction (clockwise) from the point M_0 to the point N_1. We must reckon the corresponding angle φ also clockwise, assuming φ to vary from 0 to $-\pi/2$. With such a reckoning, φ varies continuously, increasing from $-\pi/2$ to $\pi/2$ as y increases continuously from -1 to 1.

The solution of the equation $\sin \varphi = y$ belonging to the interval $[-\pi/2, \pi/2]$ is known as *arc sine*, or the *inverse sine* of y and is designated as $\arcsin y$. Thus, for instance, $\arcsin 0 = 0$, $\arcsin \frac{1}{2} = \frac{\pi}{6}$, $\arcsin \left(-\frac{\sqrt{2}}{2}\right) = -\frac{\pi}{4}$ and so on. We have established that $\arcsin y$ is a continuous function of y defined on the interval $[-1, 1]$ and increasing on this interval from $-\pi/2$ to $\pi/2$.

The solution of the equation $\sin \varphi = y$ corresponding to the point M' (Fig. 5.4) is equal to $\pi - \arcsin y$ when the angle is reckoned in the positive direction. This can be seen from Figs. 5.5 and 5.6 showing the angles for $y > 0$ and $y < 0$.

Evidently, all the solutions of the equation $\sin \varphi = y$, which correspond to the point M, are

$$\arcsin y + 2\pi k \text{ for any integer } k,$$

and the solutions which correspond to the point M' are

$$\pi - \arcsin y + 2\pi k = (2k + 1)\pi - \arcsin y$$

for any integer k.

The graph of the function $y = \sin \varphi$ (Fig. 5.7) shows all the solutions of the equation $\sin \varphi = y$. For a fixed y, the

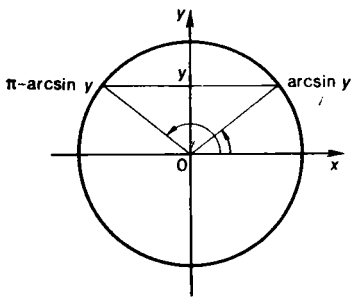


Fig. 5.5

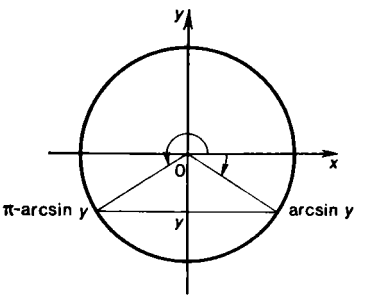


Fig. 5.6

corresponding straight line cuts the sine curve at an infinitude of points: M, M', M_1, M'_1, . . ., M'_{-1}, M_{-1}, M'_{-2}, M_{-2}, ... We designate as M, M_1, . . ., M_{-1}, M_{-2}, . . .

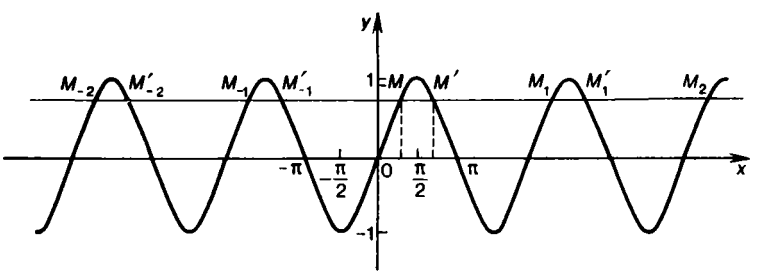


Fig. 5.7

the points on the sine curve which correspond to the point M in Fig. 5.4, and as M', M'_1, . . ., M'_{-1}, M'_{-2}, . . . the points which correspond to the point M' on a unit circle. It is clear that the solution of the equation $\sin \varphi = y$ equal

to the abscissa of the point M_k is $\arcsin y + 2\pi k$ and the abscissa of the point M'_k is $(2k + 1)\pi - \arcsin y$ for any positive or negative integer k.

The solution of the equation $\sin \varphi = y$ designated as $\arcsin y$ is the abscissa of the point M of intersection of the portion of the sine curve on the interval $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$ with a horizontal straight line corresponding to the value of y

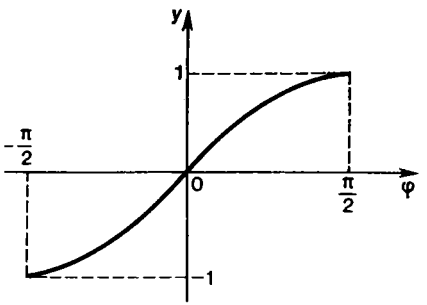


Fig. 5.8

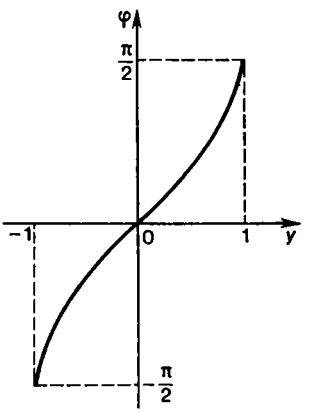


Fig. 5.9

(Fig. 5.8). We can also regard this portion of the sine curve as the graph of the function $\varphi = \arcsin y$, but in Fig. 5.8 the values of the argument y of this function are laid off along the axis of ordinates and the values of the function $\varphi = \arcsin y$ along the abscissa axis. Now if, as we usually do when constructing the graph of a function, we lay off the values of the argument along the horizontal axis and the values of the function along the vertical axis, then in Fig. 5.8 we must interchange the abscissas and ordinates of all points, and this is equivalent to the symmetry of Fig. 5.8 with respect to the bisector of the first quadrant since the points (φ, y) and (y, φ) are symmetric with respect to this bisector. We get Fig. 5.9.

The functions $y = \sin x$ for $-\frac{\pi}{2} \leqslant x \leqslant \frac{\pi}{2}$ and $y = \arcsin x$ are mutually inverse in the same sense in which an

exponential and a logarithmic function are mutually inverse or the function of raising to a power is an inverse of the extraction of a square root. Namely, from the relation $\sin \varphi = y$, $-\pi/2 \leqslant \varphi \leqslant \pi/2$, it follows that $\varphi = \arcsin y$ and from the relation $\varphi = \arcsin y$ it follows that $\sin \varphi = y$, $-\pi/2 \leqslant \varphi \leqslant \pi/2$. These relationships are similar to the following relationships: from a^x it follows that $x = \log_a y$ and from

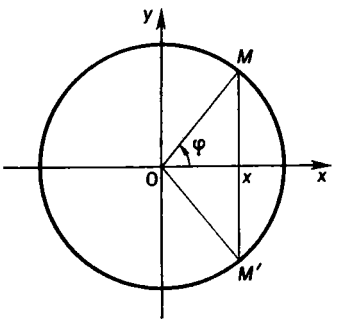


Fig. 5.10

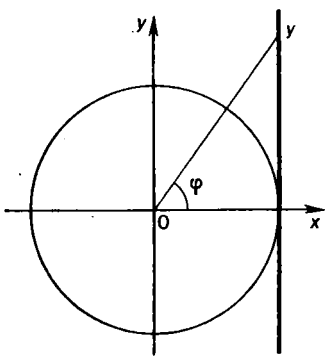


Fig. 5.11

$x = \log_a y$ it follows that $a^x = y$; it follows from $x^2 = y$, $x \geqslant 0$, that $x = \sqrt{y}$ and it follows from $x = \sqrt{y}$ that $x^2 = y$, $x \geqslant 0$.

In short, $\sin(\arcsin x) = x$ for $-1 \leqslant x \leqslant 1$ and $\arcsin(\sin x) = x$ for $-\pi/2 \leqslant x \leqslant \pi/2$ just as $\log_a a^x = x$ for all x and $a^{\log_a x} = x$ for $x > 0$ as well as $(\sqrt{x})^2 = x$ and $x = \sqrt{x}$ for $x \geqslant 0$.

Let us consider now an equation $\cos \varphi = x$. This equation has a solution for $|x| \leqslant 1$. We represent the cosine and sine functions as the coordinates of a point on a unit circle (Fig. 5.10). We can see from the figure that for $|x| \neq 1$ there are two points M and M' on a unit circle which are symmetric about the x-axis, one of which is above the x-axis and the other is below it. As x varies from 1 to -1, the angle φ corresponding to the upper point will vary continuously from 0 to π (when we make a simple reckoning in the positive direction). This angle is designated as $\arccos x$ and is called *arc cosine* of x. Thus arc cosine of x is an angle

belonging to the interval $[0, \pi]$ whose cosine is equal to x. From the definitions of the functions $\cos x$ and $\arccos x$ it follows that $\cos(\arccos x) = x$ for $|x| \leqslant 1$ and $\arccos(\cos x) = x$ for $0 \leqslant x \leqslant \pi$.

All solutions of the equation $\cos \varphi = x$ are evidently defined by the formulas

$$\varphi = \arccos x + 2\pi k, \quad \varphi = -\arccos x + 2\pi k,$$

where k is an integer. The last class of solutions corresponds to the lower point M' in Fig. 5.10.

The function $y = \tan \varphi$ (Fig. 5.11) increases continuously, passing through all real values when φ varies from $-\pi/2$ to

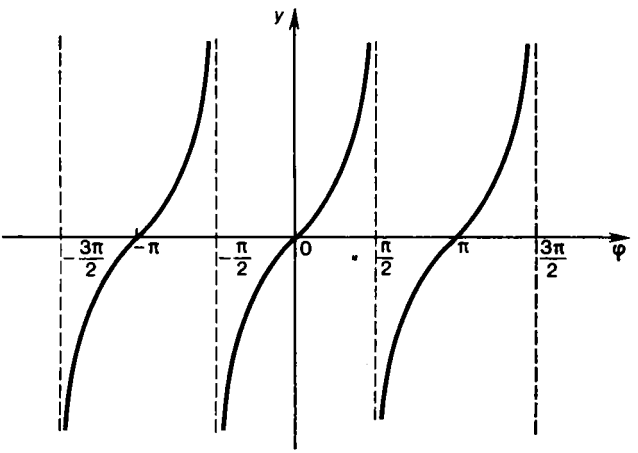


Fig. 5.12

$\pi/2$, with the exception of the endpoints. For any value of y there is one and only one value of φ in the interval $(-\pi/2, \pi/2)$ for which $\tan \varphi = y$. This value of φ is called the *arc tangent* of y and is designated as $\arctan y$. All solutions of the equation $\tan \varphi = y$ are given by the formula

$$\varphi = \arctan y + \pi k,$$

where k is an integer.

Having constructed the graph of the function $y = \tan \varphi$ (Fig. 5.12), we see that $\arctan y$ is the value of φ for the given y on the continuous portion of the graph of the tangent which passes through the origin. We can regard this portion

of the graph as the graph of the function $\varphi = \arctan y$ with the abscissa and ordinate exchanged. If we lay off the argument y of the function $\arctan y$ along the abscissa axis and the values of the function $\arctan y$ along the axis of ordinates, we get a graph which is shown in Fig. 5.13.

We designate as $\operatorname{arccot} x$ (*arc cotangent* of x) the value of φ in the interval $(0, \pi)$ for which $\cot \varphi = x$. This restriction

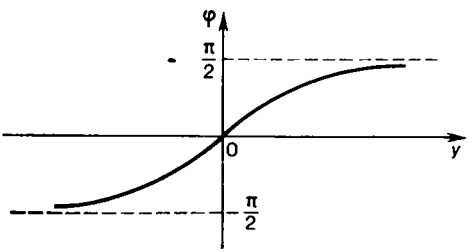


Fig. 5.13

for the angle φ is caused by the fact that for φ varying from 0 to π, $\cot \varphi$ varies continuously, decreasing throughout the number axis from ∞ to $-\infty$.

The functions arc cosine and arc cotangent are in a simple relationship with the functions arc sine and arc tangent and therefore are not frequently used. These relationships are defined by the formulas

$$\arccos x = \frac{\pi}{2} - \arcsin x, \quad \arctan x = \frac{\pi}{2} - \operatorname{arccot} x.$$

In fact, $\arccos x$ is the only value of the argument, belonging to the interval $[0, \pi]$, whose cosine is equal to x. The values of $\arcsin x$ lie in the interval $[-\pi/2, \pi/2]$ and so

$$0 \leqslant \frac{\pi}{2} - \arcsin x \leqslant \pi$$

and

$$\cos\left(\frac{\pi}{2} - \arcsin x\right) = \sin(\arcsin x) = x.$$

Therefore $\dfrac{\pi}{2} - \arcsin x = \arccos x$.

For the same reason

$$\frac{\pi}{2} - \arctan x = \operatorname{arccot} x.$$

Example 1. Find arccot (—1).

Answer. $\frac{3}{4}\pi$.

Example 2. Prove that arcsin x is an odd function.

Solution. arcsin ($-x$) is an angle, belonging to the interval $[-\pi/2, \pi/2]$, whose sine is equal to $-x$. It is equal to —arcsin x since

$$-\frac{\pi}{2} \leqslant -\arcsin x \leqslant \frac{\pi}{2},$$

$$\sin(-\arcsin x) = -\sin(\arcsin x) = -x.$$

In the same way we can prove that the function arctan x is odd.

5.11. Some Actions on Direct and Inverse Trigonometric Functions

The sine, cosine, tangent and cotangent functions can be expressed in terms of one another by means of algebraic formulas. We can therefore express the values of these functions of each of the inverse trigonometric functions in the form of algebraic expressions of the argument. Let us consider some examples.

Example 1. Express cos (arcsin x) in terms of x by means of an algebraic formula.

Solution. Note first of all that cos (arcsin x) is nonnegative since $-\pi/2 \leqslant \arcsin x \leqslant \pi/2$. Furthermore,

$$\cos(\arcsin x) = \sqrt{1 - \sin^2(\arcsin x)} = \sqrt{1 - x^2}.$$

Example 2. Prove that $\arctan x = \arcsin \frac{x}{\sqrt{1+x^2}}$.

Solution. Both sides of the equality we have to prove belong to $[-\pi/2, \pi/2]$, and so we have to make sure that their sines are equal. We have

$$\sin(\arctan x) = \tan(\arctan x)\cos(\arctan x).$$

Furthermore,

$$\cos(\arctan x) = \frac{1}{\sqrt{1+\tan^2(\arctan x)}} = \frac{1}{\sqrt{1+x^2}}$$

(we remove the minus sign before the root since $\cos(\arctan x) \geqslant (0)$. Thus

$$\sin(\arctan x) = \frac{x}{\sqrt{1+x^2}},$$

$$\sin\left(\arcsin \frac{x}{\sqrt{1+x^2}}\right) = \frac{x}{\sqrt{1+x^2}}.$$

Consequently,

$$\arctan x = \arcsin \frac{x}{\sqrt{1+x^2}},$$

and this is what we wished to prove.

Example 3. Prove that $\arctan \frac{1}{2} + \arctan \frac{1}{3} = \frac{\pi}{4}$.

Solution. From $0 < 1/2 < 1$ it follows that $0 < \arctan \frac{1}{2} < \arctan 1 = \frac{\pi}{4}$ and for the same reason $0 < \arctan \frac{1}{3} < \frac{\pi}{4}$. Hence $0 < \arctan \frac{1}{2} + \arctan \frac{1}{3} < \frac{\pi}{2}$. Furthermore,

$$\tan\left(\arctan \frac{1}{2} + \arctan \frac{1}{3}\right)$$

$$= \frac{\tan\left(\arctan \frac{1}{2}\right) + \tan\left(\arctan \frac{1}{3}\right)}{1 - \tan\left(\arctan \frac{1}{2}\right) \cdot \tan\left(\arctan \frac{1}{3}\right)} = \frac{\frac{1}{2} + \frac{1}{3}}{1 - \frac{1}{2} \cdot \frac{1}{3}} = 1$$

and, consequently, $\arctan \frac{1}{2} + \arctan \frac{1}{3} = \frac{\pi}{4}$.

Example 4.

$$\begin{aligned}\sin(2 \arctan x) &= 2 \sin(\arctan x) \cos(\arctan x) \\ &= 2 \tan(\arctan x) \cos^2(\arctan x) \\ &= \frac{2 \tan(\operatorname{arccot} x)}{1 + \tan^2(\operatorname{arccot} x)} = \frac{2x}{1+x^2}.\end{aligned}$$

Example 5.

$$\tan\left(\frac{1}{2} \arcsin x\right) = \frac{\sin(\arcsin x)}{1 + \cos(\arcsin x)} = \frac{x}{1 + \sqrt{1-x^2}}.$$

5.12. Trigonometric Equations

Equations which relate the values of trigonometric functions are called *trigonometric equations.*

Trigonometric equations can be reduced to the form $F(x) = 0$, where $F(x)$ is a periodic function which can be expressed in terms of trigonometric functions. It is clear that if x_0 is a solution of the equation $F(x) = 0$, then it yields a whole class of solutions $x_k = x_0 + kT$, where k is an integer and T is the period of $F(x)$. Therefore, if a trigonometric equation has at least one solution, then it has infinitely many solutions. However, the number of classes of solutions is usually finite. Thus the equation $\sin x = a$ has two classes of solutions for $|a| < 1$:

$$x = \arcsin a + 2k\pi, \quad x = \pi - \arcsin a + 2k\pi.$$

They merge into one class for $a = \pm 1$. The equation $\cos x = a$ also has two classes of solutions for $|a| < 1$:

$$x = \arccos a + 2k\pi, \quad x = -\arccos a + 2k\pi,$$

which merge for $a = \pm 1$. However, the equation $\tan x = a$ (the least period here is π and not 2π) has one class of solutions

$$x = \arctan a + k\pi.$$

To find all solutions of a trigonometric equation, it is sufficient to find one solution from every class. Evidently, there is one and only one solution in every class which belongs to a half-open interval whose length is equal to the period T, say, to the interval $\left[-\frac{1}{2}T, \frac{1}{2}T\right)$ or to $[0, T)$, and the number of solutions in such an interval is finite. One of the endpoints of the interval must be excluded since if one endpoint yields a solution, then the other endpoint is also a solution from the same class.

The most natural method of solution of the trigonometric equation $F(x) = 0$ consists in using algebraic means to express $F(x)$ in terms of one of the trigonometric functions. Then we solve the resulting algebraic equation and, finally, find the value of the argument from the given value of the trigonometric function. In theory it is more convenient to express $\sin x$, $\cos x$, $\tan x$ in terms of the tangent of a half-

angle, however, it often leads to a very high degree of an algebraic equation. In addition, this method may lead to a loss of the solution $x = \pi + 2k\pi$ if it exists since $\tan \frac{\pi}{2}$ is not defined.

An equation of the form

$$a_0 \sin^n x + a_1 \sin^{n-1} x \cos x + \ldots + a_n \cos^n x = 0$$

is a *homogeneous* equation. If $a_0 \neq 0$, then $\cos x = 0$ cannot yield a solution. Therefore, dividing both sides of the relation by $\cos^n x$, we arrive at an equivalent relation

$$a_0 \tan^n x + a_1 \tan^{n-1} x + \ldots + a_n = 0.$$

Now if $a_0 = 0$, we must first put the maximum possible degree of $\cos x$ before the brackets and separately equate $\cos x$ and the second factor to zero.

In many cases it is possible to reduce an equation to a homogeneous equation by inserting a factor $\sin^2 x + \cos^2 x$ equal to unity into the terms whose degrees are smaller than the maximum one. It is sometimes convenient to begin with passing to a half-angle:

$$\sin x = 2 \sin \frac{x}{2} \cos \frac{x}{2}, \quad \cos x = \cos^2 \frac{x}{2} - \sin^2 \frac{x}{2}.$$

This makes the degrees of all terms even and it becomes possible to "equalize" the degrees by inserting $\sin^2 \frac{x}{2} + \cos^2 \frac{x}{2}$ wherever necessary. Essentially, the method of reducing an equation to a homogeneous one by means of passing to a half-angle is equivalent to expressing a half-angle in terms of the tangent, but in this case we do not lose the solution $x = \pi$, provided that it exists.

Example 1. Solve the equation $2 \cos^2 x = 3 \sin x$.

Solution. It is convenient here to express the left-hand side of the equation in terms of $\sin x$. We get

$$2(1 - \sin^2 x) = 3 \sin x, \quad 2 \sin^2 x + 3 \sin x - 2 = 0,$$

$$\sin x = \frac{-3 \pm \sqrt{9+16}}{4}, \quad \sin x = \frac{1}{2}, \quad \sin x = -2.$$

The equation $\sin x = \frac{1}{2}$ yields two classes of solutions:

$x = \frac{\pi}{6} + 2k\pi$ and $x = \frac{5}{6}\pi + 2k\pi$ for any integer k. The equation $\sin x = -2$ has no solutions.

Example 2. $\sin 3x + 4 \sin x - 3 = 0$.

Solution. We must again express the left-hand side of the equation in terms of $\sin x$. We obtain

$$\begin{aligned}\sin 3x &= \sin (2x + x) = \sin 2x \cos x + \cos 2x \sin x \\ &= 2 \sin x \cos^2 x + (\cos^2 x - \sin^2 x) \sin x \\ &= 3 \sin x (1 - \sin^2 x) - \sin^3 x = 3 \sin x - 4 \sin^3 x.\end{aligned}$$

The equation reduces to

$$7 \sin x - 4 \sin^3 x - 3 = 0.$$

We set $\sin x = z$ and get a cubic equation

$$4z^3 - 7z + 3 = 0.$$

The coefficients of the initial equation have been so well chosen that the equation has an obvious root $z_1 = 1$. Consequently, the left-hand side can be divided by $z - 1$. It is easy to see that

$$4z^3 - 7z + 3 = (z - 1)(4z^2 + 4z - 3).$$

Equating the second factor to zero, we get $z_2 = 1/2$, $z_3 = -3/2$. The equation $\sin x = 1$ yields one class of solutions $\frac{\pi}{2} + 2k\pi$. The equation $\sin x = 1/2$ yields another two classes: $x = \frac{\pi}{6} + 2k\pi$ and $x = \frac{\pi}{6}\pi + 2k\pi$, where k is any integer. The equation $\sin x = -3/2$ has no solutions.

Thus the given equation has three classes of solution:

$$x = \frac{\pi}{2} + 2k\pi, \quad x = \frac{\pi}{6} + 2k\pi, \quad x = \frac{5}{6}\pi + 2k\pi,$$

where k is any integer.

Example 3. $\sin^4 x + 3 \sin x \cos^3 x = 1$.

We make the equation homogeneous by replacing 1 on the right-hand side by $(\sin^2 x + \cos^2 x)^2$. We obtain

$$\sin^4 x + 3 \sin x \cos^3 x = \sin^4 x + 2 \sin^2 x \cos^2 x + \cos^4 x.$$

Transferring all the terms to one side, we obtain

$$2 \sin^2 x \cos^2 x + \cos^4 x - 3 \sin x \cos^3 x = 0,$$

$$\cos^2 x (2 \sin^2 x + \cos^2 x - 3 \sin x \cos x) = 0.$$

The equation $\cos^2 x = 0$ yields

$$x = \frac{\pi}{2} + k\pi$$

for any integer k. The equation

$$2 \sin^2 x - 3 \sin x \cos x + \cos^2 x = 0$$

is equivalent to the equation

$$2 \tan^2 x - 3 \tan x + 1 = 0,$$

whence it follows that $\tan x = 1$ and $\tan x = 1/2$. This gives another two classes of solutions: $x = \frac{\pi}{4} + k\pi$, $x = \arctan \frac{1}{2} + k\pi$ for any integer k. The least period of the function $\sin^4 x + 3 \sin x \cos^3 x - 1$ is evidently π.

Thus the equation has three classes of solutions:

$$x = \frac{\pi}{2} + k\pi, \quad x = \frac{\pi}{4} + k\pi, \quad x = \arctan \frac{1}{2} + k\pi,$$

where k is any integer.

Example 4. $2 \sin x + \cos x = 1$.

We shall employ three techniques to solve this example.

Solution 1. $\cos x = 1 - 2 \sin x$. We square both sides of the relation and get $\cos^2 x = (1 - 2 \sin x)^2$. This equation is a consequence of the initial equation but must not necessarily be equivalent to it. Now we express its left-hand side in terms of $\sin x$ and solve the resulting equation for $\sin x$:

$$1 - \sin^2 x = 1 - 4 \sin x + 4 \sin^2 x, \quad 5 \sin^2 x - 4 \sin x = 0,$$

$$\sin x (5 \sin x - 4) = 0, \quad \sin x_1 = 0 \text{ and } \sin x_2 = 4/5.$$

Before seeking the angle from the sine, we return to the initial equation and find the cosine of x from it. We get $\cos x_1 = 1$ and $\cos x_2 = -3/5$. From $\sin x_1 = 0$ and $\cos x_1 = 1$ we get $x_1 = 2k\pi$. From $\sin x_2 = 4/5$, $\cos x_2 = -3/5$ we get $x_2 = \pi - \arcsin 4/5 + 2k\pi$ (since $\cos x_2$ is negative).

We have found two classes of solutions: $x_1 = 0 + 2k\pi$, $x_2 = \pi - \arcsin \frac{4}{5} + 2k\pi$ for any integer k. Had we found x_1 and x_2 only from the sine, we would find another two classes of solutions: $x = \pi + 2k\pi$, $x = \arcsin \frac{4}{5} + 2k\pi$. These solutions satisfy the thansformed equation but do not satisfy the initial one.

Solution 2. Recall that $\sin x = \frac{2\tan(x/2)}{1+\tan^2(x/2)}$ and $\cos x = \frac{1-\tan^2(x/2)}{1+\tan^2(x/2)}$. We have obtained an equation

$$\frac{4z+1-z^2}{1+z^2} = 1,$$

with respect to $z = \tan\frac{x}{2}$, whence it follows that $4z = 2z^2$, $z_1 = 0$, $z_2 = 2$. The equation $\tan\frac{x}{2} = 0$ yields $\frac{x}{2} = 0 + k\pi$ and $x = 0 + 2k\pi$, where k is an integer; the equation $\tan\frac{x}{2} = 2$ yields $\frac{x}{2} = \arctan 2 + k\pi$ and $x = 2\arctan 2 + 2k\pi$, where k is an integer.

It is easy to verify that $2\arctan 2 = \pi - \arcsin\frac{4}{5}$.

Solution 3. We reduce the equation to a homogeneous one by setting

$$\cos x = \cos^2\frac{x}{2} - \sin^2\frac{x}{2}, \quad \sin x = 2\sin\frac{x}{2}\cos\frac{x}{2}$$

and

$$1 = \sin^2\frac{x}{2} + \cos^2\frac{x}{2}.$$

We obtain

$$4\sin\frac{x}{2}\cos\frac{x}{2} + \cos^2\frac{x}{2} - \sin^2\frac{x}{2} = \sin^2\frac{x}{2} + \cos^2\frac{x}{2},$$

whence $2\sin^2\frac{x}{2} - 4\sin\frac{x}{2}\cos\frac{x}{2} = 0$, $2\sin\frac{x}{2}\left(\sin\frac{x}{2} - 2\cos\frac{x}{2}\right) = 0$. Equating the first factor to zero, we get $\sin\frac{x}{2} = 0$, $\frac{x}{2} = 0 + k\pi$, $x = 0 + 2k\pi$ for an integer k. To equate the second factor to zero is the same as to solve the equation $\tan\frac{x}{2} - 2 = 0$, whence $x = 2\arctan 2 + 2k\pi$ for an integer k.

Example 5. $2\sin x - \cos x = 1$.

Solution 1. $\cos x = 2\sin x - 1$, whence $\cos^2 x = (2\sin x - 1)^2$, $1 - \sin^2 x = 4\sin^2 x - 4\sin x + 1$, $5\sin^2 x - 4\sin x = 0$, $\sin x_1 = 0$, $\sin x_2 = 4/5$. We have solved, for $\sin x$,

a transformed equation which is a consequence of the initial one but is not equivalent to it. This equation proved to be exactly the same as that in Example 4. We find cos x from the initial equation:

$$\cos x_1 = -1, \quad \cos x_2 = 3/5.$$

Finally, $x_1 = \pi + 2k\pi$ for an integer k, $x_2 = \arcsin \frac{4}{5} + 2k\pi$ for an integer k. We have obtained solutions that were extraneous in Example 4 when we sought x from the sine without due account of the sign of the cosine.

Solution 2. We pass to $z = \tan \frac{x}{2}$ and obtain

$$\frac{4 \tan \frac{x}{2}}{1+\tan^2 \frac{x}{2}} - \frac{1-\tan^2 \frac{x}{2}}{1+\tan^2 \frac{x}{2}} = 1,$$

$$4z - 1 + z^2 = 1 + z^2, \quad 4z = 2, \quad z = \frac{1}{2},$$

$$x = 2 \arctan \frac{1}{2} + 2k\pi, \quad k \text{ is any integer.}$$

The second class of solutions $x = \pi + 2k\pi$ turned out to be lost. But this class, provided that it exists, is always lost when we pass to the tangent of a half-argument.

Solution 3. We replace $\sin x$ by $2 \sin \frac{x}{2} \cos \frac{x}{2}$, $\cos x$ by $\cos^2 \frac{x}{2} - \sin^2 \frac{x}{2}$, 1 by $\cos^2 \frac{x}{2} + \sin^2 \frac{x}{2}$. We get a homogeneous equation

$$4 \sin \frac{x}{2} \cos \frac{x}{2} - \cos^2 \frac{x}{2} + \sin^2 \frac{x}{2} = \cos^2 \frac{x}{2} + \sin^2 \frac{x}{2},$$

whence

$$2 \cos \frac{x}{2} \left(2 \sin \frac{x}{2} - \cos \frac{x}{2}\right) = 0.$$

Equating the first factor to zero, we get $\frac{x}{2} = \frac{\pi}{2} + k\pi$, $x = \pi + 2k\pi$ for an integer k. Equating to zero the second factor, we find that $x = 2 \arctan \frac{1}{2} + 2k\pi$, where k is any integer. This time we have not lost the class of solutions $x = \pi + 2k\pi$.

In addition to these techniques of solving trigonometric equations which reduce to the expression in terms of one function, there are other techniques based on the use of other properties of trigonometric functions and transformations of trigonometric expressions. There is a variety of techniques, but we shall discuss only several examples.

Example 6. $\sin 3x = \sin x$.

Solution 1. We have $\sin 3x = 3 \sin x - 4 \sin^3 x$. Substituting into the equation the expression for $\sin 3x$ and setting $\sin x = z$, we get an algebraic equation $3z - 4z^3 = z$, whence $z_1 = 0$, $z_2 = \frac{\pm 1}{\sqrt{2}}$, $z_3 = -\frac{1}{\sqrt{2}}$. From $\sin x = 0$ we get two classes of solutions: $x = 0 + 2k\pi$ and $x = \pi + 2k\pi$, where k is an integer. From $\sin x = 1/\sqrt{2}$ we get anoter two classes: $x = \frac{\pi}{4} + 2k\pi$ and $x = \frac{3}{4}\pi + 2k\pi$, where k is any integer. From $\sin x = -1/\sqrt{2}\pi$ we get $x = -\frac{\pi}{4} + 2k\pi$, $x = -\frac{3}{4}\pi + 2k\pi$, where k is an integer. We have got six classes all in all.

Solution 2. The sines of the arguments $3x$ and x are equal. Consequently, these arguments are connected by one of the relations $3x = x + 2k\pi$ and $3x = \pi - x + 2k\pi$, where k is any integer. The first yields $x = k\pi$ for an integral k, and this gives the first two classes of solutions, $x = 2k\pi$ and $x = \pi + 2k\pi$, for even and odd values of k. The second relation gives $x = \frac{2k+1}{4}\pi$. For $k = 0, 1, 2, 3$ we obtain $x = \frac{\pi}{4}$, $x = \frac{3}{4}\pi$, $x = \frac{5}{4}\pi = -\frac{3}{4}\pi + 2\pi$, $x = \frac{7}{4}\pi = -\frac{\pi}{4} + 2\pi$. For $k = 4, 5, \ldots, k = -1, -2, \ldots$, we shall get the same solutions with an addition of 2π, then 4π and so on, -2π, -4π and so on. Thus, for $k = 0, 1, 2, 3$ we get solutions from the other four classes.

Example 7. $a \sin x + b \cos x = c$.

Solution 1. We reduce the equation to a homogeneous one by making the substitutions $\sin x = 2 \sin \frac{x}{2} \cos \frac{x}{2}$, $\cos x = \cos^2 \frac{x}{2} - \sin^2 \frac{x}{2}$ and $1 = \cos^2 \frac{x}{2} + \sin^2 \frac{x}{2}$. We obtain

$$2a \sin \frac{x}{2} \cos \frac{x}{2} + b \left(\cos^2 \frac{x}{2} - \sin^2 \frac{x}{2} \right) = c \left(\cos^2 \frac{x}{2} + \sin^2 \frac{x}{2} \right)$$

and, after transferring to one side, we have

$$(c+b) \sin^2 \frac{x}{2} - 2a \sin \frac{x}{2} \cos \frac{x}{2} + (c-b) \cos^2 \frac{x}{2} = 0.$$

If $c+b \neq 0$, we arrive at an equivalent equation

$$(c+b) \tan^2 \frac{x}{2} - 2a \tan \frac{x}{2} + (c-b) = 0,$$

whence we have $\tan \frac{x}{2} = \frac{a \pm \sqrt{a^2+b^2-c^2}}{c+b}$, which yields two classes of solutions

$$x = 2 \arctan \frac{a \pm \sqrt{a^2+b^2-c^2}}{c+b} + 2k\pi, \quad k \text{ is an integer,}$$

only if $a^2+b^2 > c^2$. They merge into one class

$$x = 2 \arctan \frac{a}{c+b} + 2k\pi$$

if $a^2+b^2 = c^2$, and there are no solutions if $a^2+b^2 < c^2$.

Now if $b+c=0$ and $a \neq 0$, then the transformed equation is

$$\cos \frac{x}{2} \left((c-b) \cos \frac{x}{2} - 2a \sin \frac{x}{2} \right) = 0,$$

whence, equating the factors to zero, we obtain $x = \pi + 2k\pi$, $x = 2 \arctan \frac{c-b}{2a} + 2k\pi$ for an integer k.

If $b + c = 0$ and $a = 0$ but $c = -b \neq 0$, we obtain one class of solutions $x = \pi + 2k\pi$ for an integer k.

Solution 2. We divide both sides of the initial equation by $\sqrt{a^2+b^2}$ and find an angle α such that $\cos \alpha = \frac{a}{\sqrt{a^2+b^2}}$, $\sin \alpha = \frac{b}{\sqrt{a^2+b^2}}$. This is possible since the sum of the squares of the right-hand sides is equal to 1. The equation assumes the form

$$\sin x \cos \alpha + \cos x \sin \alpha = \frac{c}{\sqrt{a^2+b^2}}$$

or

$$\sin(x+\alpha)=\frac{c}{\sqrt{a^2+b^2}}.$$

This equation has solutions for $|c|\leqslant\sqrt{a^2+b^2}$ or, what is the same, for $c^2\leqslant a^2+b^2$. Under the condition that $c^2<a^2+b^2$ we have two classes of solutions:

$$x=-\alpha+\arcsin\frac{c}{\sqrt{a^2+b^2}}+2k\pi$$

and

$$x=-\alpha+\pi-\arcsin\frac{c}{\sqrt{a^2+b^2}}+2k\pi,\quad k \text{ is an integer.}$$

They merge into one class for $c=\sqrt{a^2+b^2}$ and for $c=-\sqrt{a^2+b^2}$. In the first case $x=-\alpha+\arcsin 1+2k\pi=-\alpha+\frac{\pi}{2}+2k\pi$. In the second case $x=-\alpha+\arcsin(-1)+2k\pi=-\alpha-\frac{\pi}{2}+2k\pi$.

Example 8. $\sin x+\sin 2x+\sin 3x+\sin 4x+\sin 5x=0$.

Solution. We use the technique of "contracting" the sums of the values of trigonometric functions described in 5.6. We have

$$2\sin\frac{x}{2}(\sin x+\sin 2x+\sin 3x+\sin 4x+\sin 5x)$$

$$=2\sin\frac{x}{2}\sin x+2\sin\frac{x}{2}\sin 2x+2\sin\frac{x}{2}\sin 3x$$

$$+2\sin\frac{x}{2}\sin 4x+2\sin\frac{x}{2}\sin 5x=\cos\frac{x}{2}-\cos\frac{3}{2}x$$

$$+\cos\frac{3}{2}x-\cos\frac{5}{2}x+\cos\frac{5}{2}x-\cos\frac{7}{2}x+\cos\frac{7}{2}x-\cos\frac{9}{2}x$$

$$+\cos\frac{9}{2}x-\cos\frac{11}{2}x=\cos\frac{1}{2}x-\cos\frac{11}{2}x=2\sin 3x\cdot\sin\frac{5}{2}x,$$

and so the equation reduces to $\dfrac{\sin 3x\cdot\sin\frac{5}{2}x}{\sin\frac{x}{2}}=0.$

Equating the numerator to zero, we get $\sin 3x=0$ or $\sin\frac{5}{2}x=0$, whence $3x=k\pi$ or $\frac{5}{2}x=k\pi$ and $x=\frac{k\pi}{3}$ and $x=\frac{2k\pi}{5}$, where k is an integer.

The period of the left-hand side of the initial equation is equal to 2π. To obtain representatives of all classes of solutions, we must set $k = 0, 1, 2, 3, 4, 5$ in the first case and $k = 0, 1, 2, 3, 4$ in the second case. In both cases, for $k = 0$, we get the same class containing $x = 0$. Although for $x = 0$ the denominator of the transformed equation vanishes so that $x = 0$ is not a root of the transformed equation, $x = 0$ is a root of the initial equation. As to the solution $\frac{\pi}{3}, \frac{2\pi}{3}, \pi, \frac{4}{3}\pi, \frac{5}{3}\pi, \frac{2}{5}\pi, \frac{4}{5}\pi, \frac{6}{5}\pi, \frac{8}{5}\pi$, they belong to different classes pairwise, and so the initial equation has ten classes of solutions.

5.13. Solving Basic Trigonometric Inequalities

In this section we shall denote the argument of trigonometric functions by the letter φ because we shall use the letters x

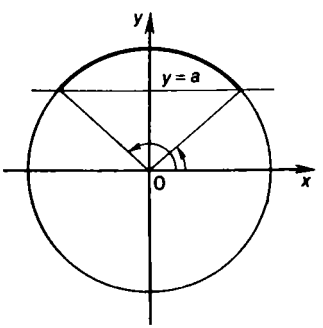


Fig. 5.14

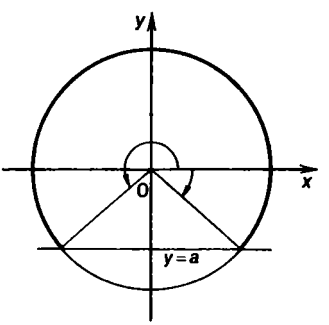


Fig. 5.15

and y for the coordinates of points in the plane with a circle of radius 1 depicted in it.

We begin with the inequality $\sin \varphi \geqslant a$, where $0 \leqslant a < 1$. The points on a unit circle corresponding to the arguments φ lie above the straight line $y = a$ and on the straight line itself (Fig. 5.14). The arguments evidently satisfy the inequalities $\arcsin a \leqslant \varphi \leqslant \pi - \arcsin a$ and, in a more general form, the inequalities $\arcsin a + 2k\pi \leqslant \varphi \leqslant \pi - \arcsin a + 2k\pi$. When we solve the inequality $\sin \varphi > a$ and describe its solutions, we must exclude the equality signs.

For $-1 < a < 0$, the arguments which yield solutions of the inequality $\sin \varphi \geqslant a$, fill in the arc of the unit circle lying above the straight line $y = a$ (Fig. 5.15). This arc is longer than the semicircle. It is again characterized by the

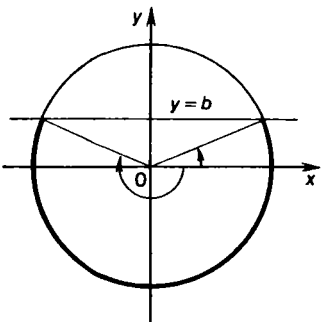


Fig. 5.16

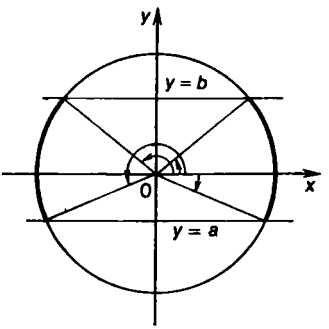


Fig. 5.17

inequalities $\arcsin a \leqslant \varphi \leqslant \pi - \arcsin a$ and, in a more general form, by the inequalities $\arcsin a + 2k\pi \leqslant \varphi \leqslant \pi - \arcsin a + 2k\pi$.

The inequality $\sin \varphi \leqslant b$ can be solved in the same way. It is described by the inequalities $-\pi - \arcsin b \leqslant \varphi \leqslant \arcsin b$ and, in a more general form, by the inequalities $(2k-1)\pi - \arcsin b \leqslant \varphi \leqslant \arcsin b + 2k\pi$ (Fig. 5.16).

The solutions of the two-sided inequality $a \leqslant \sin \varphi \leqslant b$, where $-1 \leqslant a < b \leqslant 1$, correspond to the points of the unit circle within the strip between the straight lines $y = a$ and $y = b$. These points fill in two arcs (which merge for $a = -1$ or for $b = 1$), their endpoints inclusive (Fig. 5.17). The corresponding arguments are characterized by two two-sided inequalities

$$\arcsin a \leqslant \varphi \leqslant \arcsin b,$$

$$\pi - \arcsin b \leqslant \varphi \leqslant \pi - \arcsin a.$$

We can add $2k\pi$ to the arguments for integral k and obtain the inequalities

$$\arcsin a + 2k\pi \leqslant \varphi \leqslant \arcsin b + 2k\pi,$$

$$(2k+1)\pi - \arcsin b \leqslant \varphi \leqslant (2k+1)\pi - \arcsin a.$$

If the equality signs are excluded from the initial inequality for a or for b, then we must exclude the corresponding equality signs in the description of the solutions. The solutions can be illustrated by means of the graph of the function $y = \sin \varphi$ (Fig. 5.18).

The solutions described by the inequalities

$$\arcsin a + 2k\pi \leqslant \varphi \leqslant \arcsin b + 2k\pi$$

correspond to two sine curves cut out by the strip between the straight lines $y = a$ and $y = b$ on the ascending parts of the sine curve, and the solutions

$$(2k + 1)\,\pi - \arcsin b \leqslant \varphi \leqslant (2k + 1)\,\pi - \arcsin a$$

correspond to similar arcs of the descending parts of the sine curve.

The solutions of the inequalities $\cos \varphi \geqslant a$, $\cos \varphi \leqslant b$ and $a \leqslant \cos \varphi \leqslant b$ are similar. The points on the unit circle

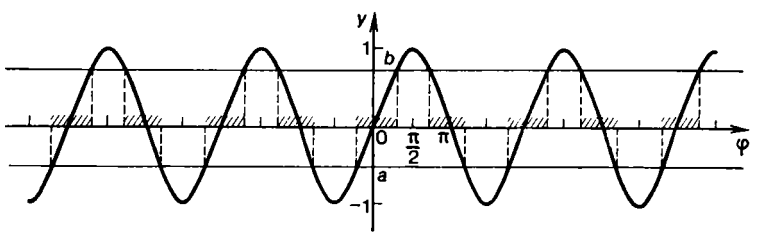


Fig. 5.18

corresponding to the solutions of the first inequality lie on the right of the straight line $x = a$ and the solutions themselves are described by the formula

$$-\arccos a + 2k\pi \leqslant \varphi \leqslant \arccos a + 2k\pi$$

(Fig. 5.19). The points corresponding to the solutions of the second inequality lie on the left of the straight line $x = b$ and the solutions are characterized by the inequalities

$$\arccos b + 2k\pi \leqslant \varphi \leqslant 2\pi - \arccos b + 2k\pi.$$

The points corresponding to the solutions of the two-sided inequality lie on the arcs cut out by the strip between the straight lines $x = a$ and $x = b$. The solutions are described

by the inequalities

$$\arccos b + 2k\pi \leqslant \varphi \leqslant \arccos a + 2k\pi,$$
$$-\arccos a + 2k\pi \leqslant \varphi \leqslant -\arccos b + 2k\pi.$$

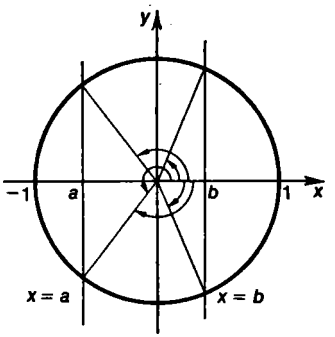


Fig. 5.19

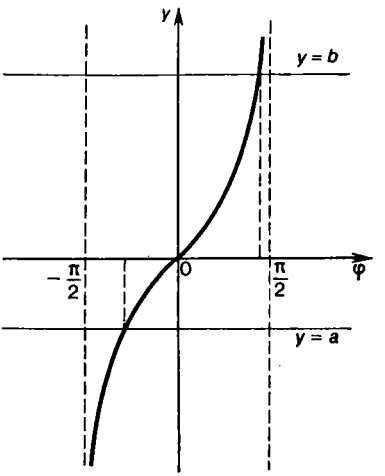


Fig. 5.20

Expressing arc cosine in terms of the arc sine, we get inequalities

$$-\frac{\pi}{2} + \arcsin a + 2k\pi \leqslant \varphi \leqslant \frac{\pi}{2} - \arcsin a + 2k\pi,$$

to solve the inequality $\cos \varphi \geqslant a$, inequalities

$$\frac{\pi}{2} - \arcsin b + 2k\pi \leqslant \varphi \leqslant \frac{3}{2}\pi + \arcsin b + 2k\pi,$$

to solve the inequality $\cos \varphi \leqslant b$, and inequalities

$$-\arcsin b + \frac{\pi}{2} + 2k\pi \leqslant \varphi \leqslant \frac{\pi}{2} - \arcsin a + 2k\pi$$

and

$$\arcsin a - \frac{\pi}{2} + 2k\pi \leqslant \varphi \leqslant \arcsin b - \frac{\pi}{2} + 2k\pi$$

to solve the inequality $a \leqslant \cos \varphi \leqslant b$.

We should have arrived at these formulas if we used the fact that $\cos \varphi = \sin (\varphi + \frac{\pi}{2})$ and applied the corresponding inequalities for the sine.

The tangent function has a period π, is continuous and increases on the interval $(-\pi/2, \pi/2)$, $\tan \varphi \to +\infty$ for $\varphi \to \pi/2$ and $\varphi < \frac{\pi}{2}$, $\tan \varphi \to -\infty$ for $\varphi \to -\pi/2$ and $\varphi > -\pi/2$. Therefore, for $-\pi/2 < \varphi < \pi/2$, the solution of the inequality $\tan \varphi \geqslant a$ is given by the formula $\arctan a \leqslant \varphi < \pi/2$ (Fig. 5.20), and the formula $\arctan a + k\pi \leqslant \varphi < \frac{\pi}{2} + k\pi$ contains all solutions of this inequality. Correspondingly, for $-\pi/2 < \varphi < \pi/2$, the solutions of the inequality $\tan \varphi \leqslant b$ are given by the formula $-\pi/2 < \varphi \leqslant \arctan b$. Finally, the solutions of the two-sided inequality $a \leqslant \tan \varphi \leqslant b$ are described by the formula

$$\arctan a + k\pi \leqslant \varphi \leqslant \arctan b + k\pi.$$

5.14. Trigonometric Inequalities of a More General Form

Let us consider inequalities of the form $F(x) \geqslant 0$, where $F(x)$ is a rational function of trigonometric functions. To solve inequalities of this kind, we can use the following two techniques. First, we can express $F(x)$ in terms of one of the trigonometric functions, say, the tangent of a half-angle. Assuming this function to be a new unknown, we get an algebraic inequality with respect to this unknown. To solve this inequality, we can employ the techniques studied in algebra course. As a result, we get one interval or several intervals for the new unknown. On the basis of the investigation of the basic trigonometric inequalities presented in the preceding section, we can describe the solutions of the initial inequality.

The second technique is based on the following property of continuous functions which seems to be obvious: *on the interval in which a function is continuous and does not vanish, its values retain sign.* If we follow this course when investigating the inequality $F(x) \geqslant 0$, we must find all points of discontinuity of $F(x)$ and all values of x for which $F(x) = 0$, arrange all of those points in the order of increase of their

abscissas and consider the sign of $F(x)$ on each interval between the marked points. For a periodic function $F(x)$, it is sufficient to do this on the interval equal to the period in length.

Example 1. Solve the inequality $2\cos^2 x > 3\sin x$.

Solution. Expressing the cosine in terms of the sine and transferring all the terms to one side, we arrive at an equivalent inequality $2 - 2\sin^2 x - 3\sin x > 0$. Setting $\sin x = z$, we get an algebraic inequality $-2z^2 - 3z + 2 > 0$. We know that a quadratic trinomial with a negative leading coefficient assumes positive values in the interval between its roots. The roots $z_1 = -2$ and $z_2 = 1/2$ are such roots. Thus the solutions of the algebraic inequality are described by the inequality $-2 < z < 1/2$. Now we have only to solve the simple trigonometric inequality $-2 < \sin x < 1/2$. The inequality $-2 < \sin x$ is always satisfied and should not be taken into account. The solutions of the inequality $\sin x < \frac{1}{2}$ are described by the formula $\frac{5\pi}{6} < x < \frac{\pi}{6} + 2\pi$ and, in the whole, by the formulas $\frac{5\pi}{6} + 2k\pi < x < \frac{\pi}{6} + 2(k+1)\pi$, where k is any integer.

Example 2. $\sin 3x \geqslant \sin x$.

Solution 1. We express $\sin 3x$ in terms of $\sin x$:

$$\sin 3x = 3\sin x - 4\sin^3 x.$$

The initial inequality assumes the form

$$2\sin x - 4\sin^3 x \geqslant 0.$$

Setting $\sin x = z$, we get $2z(1-2z^2) \geqslant 0$, $4z\left(\frac{1}{\sqrt{2}} - z\right) \times \left(\frac{1}{\sqrt{2}} + z\right) \geqslant 0$. The roots of the left-hand side are $-1/\sqrt{2}$, 0, $1/\sqrt{2}$. The left-hand side is nonnegative on the intervals $(-\infty, -1/\sqrt{2})$ and $[0, 1/\sqrt{2}]$. The first interval generates an inequality for the sine: $\sin x \leqslant -\frac{1}{\sqrt{2}}$, whence $\frac{5}{4}\pi \leqslant x \leqslant \frac{7}{4}\pi$. The second interval generates an inequality $0 \leqslant \sin x \leqslant \frac{1}{\sqrt{2}}$, whence we obtain two intervals, $0 \leqslant x \leqslant \frac{\pi}{4}$

and $\frac{3\pi}{4} \leqslant x \leqslant \pi$, for x. We have described the solutions which are in the interval $[0, 2\pi)$. The other solutions can be obtained by adding $2k\pi$, for an integer k, to the boundaries of x. On a unit circle, the set of solutions fill in three arcs shown in Fig. 5.21.

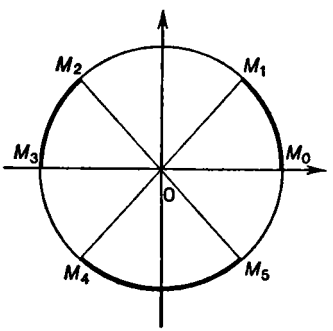


Fig. 5.21

Solution 2. The equation $\sin 3x = \sin x$ was solved in Example 6 in 5.12. Its six roots (in the limits of one period) divide the unit circle into six arcs. To find out whether the arc defines the solution or not it is sufficient to determine the sign of $F(x) = \sin 3x - \sin x$ at least at one point of the arc. We take $x = \frac{\pi}{2}$ on the arc M_0M_1. Then $F\left(\frac{\pi}{6}\right) = 1 - \frac{1}{2} = \frac{1}{2} > 0$. Consequently, the arc M_0M_1 enters into the geometric representation of the set of solutions. On the arc M_1M_2

$$x = \frac{\pi}{2}, \quad F(x) = -1 - 1 = -2 < 0.$$

The arc M_1M_2 does not enter into the geometric representation of the set of the solutions. On the arc M_2M_3

$$x = \frac{5\pi}{6}, \quad F(x) = 1 - \frac{1}{2} = \frac{1}{2} > 0.$$

The arc M_2M_3 enters into the geometric representation of the set of the solutions. On the arc M_3M_4

$$x = \frac{7\pi}{6}, \quad F(x) = -1 + \frac{1}{2} = -\frac{1}{2} < 0.$$

The arc M_3M_4 does not enter into the geometric representation of the set of the solutions. On the arc M_4M_5

$$x = -\pi/2, \quad F(-\pi/2) = 1 - (-1) > 0.$$

The arc M_4M_5 enters into the geometric representation of the set of the solutions. Finally, on the arc M_5M_0

$$x = -\frac{\pi}{6}, \quad F(x) = -1 + \frac{1}{2} = -\frac{1}{2} < 0.$$

The arc M_5M_0 does not enter into the geometric representation of the set of the solutions.

◀ **Example 3.** Solve the inequality $\tan 2x - 3 \tan x > 0$.

Solution. The least period of the function $F(x) = \tan 2x - 3 \tan x$ is equal to π and so it is sufficient to investigate the inequality on the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$. The endpoints of this interval are points of discontinuity for $\tan x$ and, consequently, for $F(x)$ as well. In addition, the points of discontinuity of the function $\tan 2x$, i.e. $-\frac{\pi}{4}$ and $\frac{\pi}{4}$ (in the limits of the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$) are points of discontinuity for $F(x)$. Let us investigate the behaviour of the function $F(x)$ in the vicinity of the points of discontinuity. For this purpose we consider the table

x	$\tan x$	$\tan 2x$	$F(x)$
$x \to -\frac{\pi}{2}$ from the right	$\to -\infty$	$\to 0$	$\to +\infty$
$x \to -\frac{\pi}{4}$ from the left	$\to -1$	$\to +\infty$	$\to +\infty$
$x \to -\frac{\pi}{4}$ from the right	$\to -1$	$\to -\infty$	$\to -\infty$
$x \to \frac{\pi}{4}$ from the left	$\to 1$	$\to +\infty$	$\to +\infty$
$x \to \frac{\pi}{4}$ from the right	$\to 1$	$\to -\infty$	$\to -\infty$
$x \to \frac{\pi}{2}$ from the left	$\to +\infty$	$\to 0$	$\to -\infty$

Thus, in the vicinity of the points of discontinuity the graph has the form shown in Fig. 5.22.

It is easy to solve the equation $\tan 2x - 3 \tan x = 0$. It is equivalent to the equation $3 \tan^3 x - \tan x = 0$, whence $\tan x_1 = -1/\sqrt{3}$, $\tan x_2 = 0$ and $\tan x_3 = 1/\sqrt{3}$ and, within the interval $(-\pi/2, \pi/2)$, $x_1 = -\pi/6$, $x_2 = 0$ and $x_3 = \pi/6$. Together with the points of discontinuity,

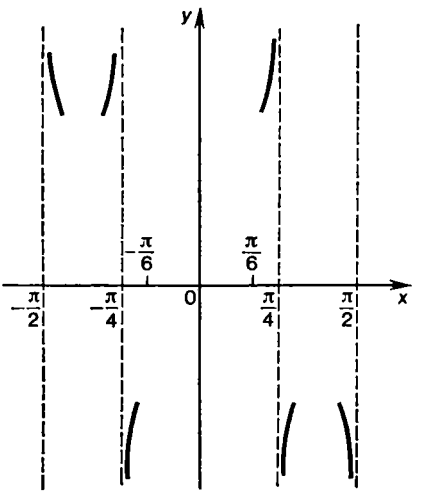


Fig. 5.22

these points divide the interval $(-\pi/2, \pi/2)$ into six subintervals on each of which $F(x)$ retains sign. From the investigation (Fig. 5.22) of the behaviour of $F(x)$ on the interval $(-\pi/2, -\pi/4)$ we infer that $F(x)$ is positive on it. On the interval $(-\pi/4, -\pi/6)$ the function $F(x)$ is negative since it is negative somewhat to the right of the point $-\pi/4$. For the same reason $F(x)$ is positive on $(\pi/6, \pi/4)$ and negative on $(\pi/4, \pi/2)$. It remains to consider the intervals $(-\pi/6, 0)$ and $(0, \pi/6)$ adjoining the point 0. We have

$$F(x) = \tan 2x - 3 \tan x = \frac{2 \tan x}{1 - \tan^2 x} - 3 \tan x$$
$$= \tan x \left(\frac{2}{1 - \tan^2 x} - 3 \right).$$

The second factor tends to -1 as $x \to 0$. The first factor, $\tan x$, changes sign from minus to plus when x increases and

passes through 0. Consequently, $F(x)$ changes sign from plus to minus, i.e. $F(x)$ is positive on the left of the point $x = 0$ and, consequently, throughout the interval $(-\pi/6, 0)$, and negative on the right of this point on the interval $(0, \pi/6)$.

Thus, on the interval $(-\pi/2, \pi/2)$ the solutions of the inequality are given by the points of the intervals $(-\pi/2, -\pi/4)$, $(-\pi/6, 0)$ and $(\pi/6, \pi/4)$. All the solutions can be obtained from these solutions by addition of $k\pi$ to the endpoints of the intervals for any integer k.

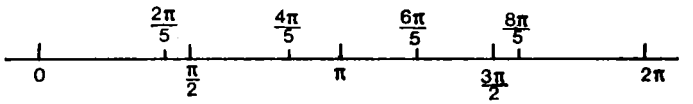


Fig. 5.23

Example 4.

$$F(x) = \sin x + \sin 2x + \sin 3x + \sin 4x > 0.$$

Solution. The left-hand side of $F(x)$ has a period 2π. For $x = 0$ and $x = 2\pi$ it vanishes. As we did in Example 7 from 5.12, we reduce the left-hand side of $F(x)$ to the form

$$F(x) = \frac{\sin 2x \sin \frac{5}{2} x}{\sin \frac{x}{2}}.$$

On the interval $(0, 2\pi)$ the denominator $\sin \frac{x}{2}$ is positive. The numerator vanishes in the roots $\pi/2$, π, $3\pi/2$ of the factor $\sin 2x$ and in the roots $2\pi/5$, $4\pi/5$, $6\pi/5$, $8\pi/5$ of the factor $\sin \frac{5}{2} x$ on the interval $[0, 2\pi]$, except for the points $x = 0$ and $x = 2\pi$. When passing through each of these roots, the corresponding factor of the numerator and, together with it, the function $F(x)$, change sign. The roots arranged in an increasing order divide the interval $(0, 2\pi)$ into eight parts (Fig. 5.23). On adjacent intervals, $F(x)$ has values of opposite signs. Evidently, $F(x) > 0$ for a sufficiently small positive x. Therefore, $F(x)$ is positive on the first, the third, the fifth, and the seventh interval and negative on the

second, the fourth, the sixth, and the eighth interval. Thus, for $0 \leqslant x < 2\pi$, the solutions of the inequality $F(x) > 0$ are given by the values of x from the intervals

$$\left(0, \frac{2}{5}\pi\right), \quad \left(\frac{\pi}{2}, \frac{4}{5}\pi\right), \quad \left(\pi, \frac{6}{5}\pi\right) \text{ and } \left(\frac{3}{2}\pi, \frac{8}{5}\pi\right). \blacktriangleright$$

◀ 5.15. Sample Proofs of Inequalities Involving Trigonometric Expressions

Sometimes, when the theory of trigonometric functions is applied to various problems, it turns out to be necessary to prove some inequalities.

When trigonometric inequalities are proved, some well-known algebraic inequalities are used together with specific trigonometric inequalities such, for instance, as $-1 \leqslant \sin x \leqslant 1$, $-1 \leqslant \cos x \leqslant 1$. When carrying out the proof, we have to use the properties of trigonometric functions, in particular, the relationship between the values of trigonometric functions. However, it is impossible to give any general rules which can be used to prove inequalities including trigonometric expressions, and so we shall only consider some examples which will demonstrate various ideas involved in the proof.

Example 1. Prove that $\cos^2 \varphi + \tan^2 \varphi \geqslant 2 \sin \varphi$.

Proof. $\cos^2 \varphi + \tan^2 \varphi - 2 \sin \varphi = \cos^2 \varphi + \tan^2 \varphi - 2 \cos \varphi \tan \varphi = (\cos \varphi - \tan \varphi)^2 \geqslant 0$, and this is what we wished to prove.

The inequality we have proved becomes trivial when $\sin \varphi < 0$. It is easy to prove a stronger inequality

$$\cos^2 \varphi + \tan^2 \varphi \geqslant 2 \mid \sin \varphi \mid.$$

Example 2. Prove that $\mid a \cos x + b \sin x \mid \leqslant \sqrt{a^2 + b^2}$.

Proof. We set $a = r \cos \alpha$, $b = r \sin \alpha$. Then $a^2 + b^2 = r^2 (\cos^2 \alpha + \sin^2 \alpha) = r^2$, $\cos \alpha = \frac{a}{r}$, $\sin \alpha = \frac{b}{r}$. The argument α which satisfies the last requirements, exists since

$$\left(\frac{a}{r}\right)^2 + \left(\frac{b}{r}\right)^2 = \frac{a^2 + b^2}{r^2} = 1.$$

Furthermore,

$$\begin{aligned} a \sin \varphi + b \cos \varphi &= r \cos \alpha \sin \varphi + r \sin \alpha \cos \varphi \\ &= r \sin (\varphi + \alpha) \end{aligned}$$

and

$$|a \sin \varphi + b \cos \varphi| = r |\sin(\varphi + \alpha)| \leqslant r = \sqrt{a^2 + b^2}.$$

Example 3. Prove that $4 \cos \varphi + 3 + \cos 2\varphi \geqslant 0$.

Proof.

$$\begin{aligned} 4 \cos \varphi + 3 + \cos 2\varphi &= 4 \cos \varphi + 3 + 2 \cos^2 \varphi - 1 \\ &= 2 (\cos^2 \varphi + 2 \cos \varphi + 1) \\ &= 2 (\cos \varphi + 1)^2 \geqslant 0. \end{aligned}$$

Example 4. Prove that if $\sin^2 x + \sin^2 y < 1$, then $|\tan x \tan y| < 1$.

We give two kinds of proof.

Proof 1. Consider

$$\begin{aligned} (\tan x \tan y)^2 - 1 &= \frac{\sin^2 x \sin^2 y}{\cos^2 x \cos^2 y} - 1 \\ &= \frac{\sin^2 x \sin^2 y - (1 - \sin^2 x)(1 - \sin^2 y)}{\cos^2 x \cos^2 y} \\ &= \frac{\sin^2 x \sin^2 y - 1 + \sin^2 x + \sin^2 y - \sin^2 x \sin^2 y}{\cos^2 x \cos^2 y} \\ &= \frac{\sin^2 x + \sin^2 y - 1}{\cos^2 x \cos^2 y} < 0, \end{aligned}$$

whence $(\tan x \tan y)^2 < 1$, $|\tan x \tan y| < 1$.

Proof 2. From the inequality $\sin^2 x + \sin^2 y < 1$ it follows that $\sin^2 x < 1 - \sin^2 y = \cos^2 y$ and $\sin^2 y < \cos^2 x$, whence

$$(\tan x \tan y)^2 = \frac{\sin^2 x \sin^2 y}{\cos^2 x \cos^2 y} = \frac{\sin^2 x}{\cos^2 y} \cdot \frac{\sin^2 y}{\cos^2 x} < 1.$$

Example 5. Prove that if $|a| \leqslant 1$ and $|b| \leqslant 1$, then $|a\sqrt{1 - b^2} + b\sqrt{1 - a^2}| \leqslant 1$.

Proof. We set $a = \cos \alpha$, $\sqrt{1 - a^2} = \sin \alpha$. Such an α exists since $a^2 + (\sqrt{1 - a^2})^2 = 1$. Similarly, $b = \cos \beta$, $\sqrt{1 - b^2} = \sin \beta$. Then $a\sqrt{1 - b^2} + b\sqrt{1 - a^2} = \cos \alpha \sin \beta + \sin \alpha \cos \beta = \sin(\alpha + \beta)$ and

$$|a\sqrt{1 - b^2} + b\sqrt{1 - a^2}| = |\sin(\alpha + \beta)| \leqslant 1.$$

It would be technically more difficult to give a direct algebraic proof of this inequality. ▶

5.16. Differentials and Derivatives of Trigonometric Functions

To make the representation of the differentials and derivatives of the cosine and sine functions more graphic, let us sketch them on a circle of radius 1. We designate the argument of these functions as φ and shall consider it, for simplicity, to be an acute angle (Fig. 5.24). It is equal to the

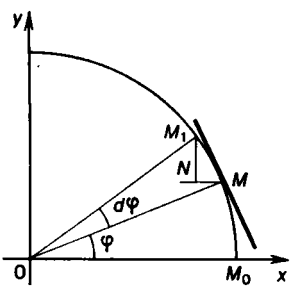


Fig. 5.24

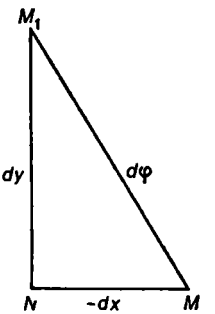


Fig. 5.25

length of the arc M_0M. We draw a tangent to the circle at the point M. It is perpendicular to the radius OM.

Let $d\varphi$ be an infinitesimal increment of the argument. It is the length of the infinitesimal arc M_1M which almost touches a segment of the tangent. We draw a vertical line through the point M_1 and a horizontal line through the point M. Let them meet at a point N. We sketch the triangle MM_1N strongly magnified (Fig. 5.25), identifying the arc MM_1 with a segment of the tangent and the increments of the coordinates $x = \cos\varphi$ and $y = \sin\varphi$ with their differentials. The angle at the vertex M_1 (after the arc is replaced by the segment of the tangent) is equal to the angle φ since the sides M_1N and M_1M of this angle are perpendicular to the sides OM_0 and OM of the angle φ.

Thus,

$$dx = d\cos\varphi = -MN = -MM_1\sin\varphi = -\sin\varphi\ d\varphi,$$

$$dy = d\sin\varphi = M_1N = MM_1\cos\varphi = \cos\varphi\ d\varphi.$$

Hence

$$(\cos \varphi)' = -\sin \varphi, \quad (\sin \varphi)' = \cos \varphi.$$

◀ We can replace these graphic but not very formal arguments by more rigorous arguments based on the limiting relation $\lim\limits_{h \to 0} \frac{\sin h}{h} = 1$. It follows from this relation that $\lim\limits_{h \to 0} \frac{1-\cos h}{h} = 0$. Indeed,

$$\frac{1-\cos h}{h} = \frac{(1-\cos h)(1+\cos h)}{h(1+\cos h)} = \frac{\sin^2 h}{h} \cdot \frac{1}{1+\cos h}$$

$$= \frac{\sin h}{h} \cdot \frac{1}{1+\cos h} \cdot \sin h,$$

the first factor tends to 1 as $h \to 0$, the second factor tends to 1/2 and the third to zero, and so we have $\lim\limits_{h \to 0} \frac{1-\cos h}{h} = 0$.

Furthermore,

$$(\cos \varphi)' = \lim_{\Delta\varphi \to 0} \frac{\cos(\varphi+\Delta\varphi) - \cos \varphi}{\Delta\varphi}$$

$$= \lim_{\Delta\varphi \to 0} \frac{\cos \varphi \cos \Delta\varphi - \sin \varphi \sin \Delta\varphi - \cos \varphi}{\Delta\varphi}$$

$$= \lim_{\Delta\varphi \to 0} \left(-\sin \varphi \frac{\sin \Delta\varphi}{\Delta\varphi} - \cos \varphi \frac{1-\cos \Delta\varphi}{\Delta\varphi} \right)$$

$$= -\sin \varphi.$$

Thus $(\cos \varphi)' = -\sin \varphi$. Similarly,

$$(\sin \varphi)' = \lim_{\Delta\varphi \to 0} \frac{\sin(\varphi+\Delta\varphi) - \sin \varphi}{\Delta\varphi}$$

$$= \lim_{\Delta\varphi \to 0} \frac{\sin \varphi \cos \Delta\varphi + \cos \varphi \sin \Delta\varphi - \sin \varphi}{\Delta\varphi}$$

$$= \lim_{\Delta\varphi \to 0} \left(\cos \varphi \frac{\sin \Delta\varphi}{\Delta\varphi} - \sin \varphi \frac{1-\cos \Delta\varphi}{\Delta\varphi} \right) = \cos \varphi,$$

and so $(\sin \varphi)' = \cos \varphi$. ▶

We can find the derivative of the tangent using the rule of calculation of the derivative of a fraction:

$$(\tan \varphi)' = \left(\frac{\sin \varphi}{\cos \varphi}\right)' = \frac{\cos \varphi (\sin \varphi)' - \sin \varphi (\cos \varphi)'}{\cos^2 \varphi} = \frac{\cos^2 \varphi + \sin^2 \varphi}{\cos^2 \varphi} = \frac{1}{\cos^2 \varphi}.$$

Thus $(\tan \varphi)' = \frac{1}{\cos^2 \varphi}$ and $d \tan \varphi = \frac{d\varphi}{\cos^2 \varphi}$.

For the cotangent we have

$$(\cot \varphi)' = \left(\frac{\cos \varphi}{\sin \varphi}\right)' = \frac{\sin \varphi (\cos \varphi)' - \cos \varphi (\sin \varphi)'}{\sin^2 \varphi} = \frac{-\sin^2 \varphi - \cos^2 \varphi}{\sin^2 \varphi} = \frac{-1}{\sin^2 \varphi},$$

and so $(\cot \varphi)' = -\frac{1}{\sin^2 \varphi}$ and $d \cot \varphi = -\frac{d\varphi}{\sin^2 \varphi}$.

Let us consider some examples showing how to calculate the derivatives and differentials of expressions dependent on trigonometric functions.

Example 1. Given $y = (\sin \varphi + \cos \varphi)^2$, find dy and y'.

Solution.

$$dy = 2 (\sin \varphi + \cos \varphi)\, d (\sin \varphi + \cos \varphi)$$
$$= 2 (\sin \varphi + \cos \varphi) (\cos \varphi - \sin \varphi)\, d\varphi$$
$$= 2 (\cos^2 \varphi - \sin^2 \varphi)\, d\varphi = 2 \cos 2\varphi d\varphi;$$
$$y' = 2 \cos 2\varphi.$$

To put it otherwise,

$$y = \sin^2 \varphi + 2 \sin \varphi \cos \varphi + \cos^2 \varphi = 1 + \sin 2\varphi,$$
$$dy = d \sin 2\varphi = \cos 2\varphi d(2\varphi) = 2 \cos 2\varphi d\varphi.$$

Example 2. Given $y = \frac{1 - \tan x}{1 + \tan x}$, find y'.

Solution.

$$y' = \frac{(1+\tan x)(1-\tan x)' - (1-\tan x)(1+\tan x)'}{(1+\tan x)^2}$$

$$= \frac{-(1+\tan x)\dfrac{1}{\cos^2 x} - (1-\tan x)\dfrac{1}{\cos^2 x}}{(1+\tan x)^2}$$

$$= \frac{-2}{\cos^2 x\,(1+\tan x)^2}.$$

We can transform the answer as follows:

$$y' = -\frac{2}{\cos^2 x\,(1+\tan x)^2} = -\frac{2}{(\cos x+\cos x\tan x)^2}$$

$$= \frac{-2}{(\cos x+\sin x)^2}$$

$$= -\frac{2}{\cos^2 x+2\cos x\sin x+\sin^2 x} = -\frac{2}{1+\sin 2x}.$$

Example 3. Given $y = \sin^4 x + \cos^4 x$, find y'.

Solution.

$$y' = 4\sin^3 x\cos x + 4\cos^3 x\,(-\sin x)$$

$$= 4\sin x\cos x\,(\sin^2 x - \cos^2 x)$$

$$= -2\sin 2x\cos 2x = -\sin 4x.$$

To put it otherwise,

$$y = (\sin^2 x+\cos^2 x)^2 - 2\sin^2 x\cos^2 x = 1 - \frac{1}{2}\sin^2 2x,$$

$$y' = -\frac{1}{2}\cdot 2\sin 2x\cos 2x\cdot 2 = -\sin 4x.$$

Let us consider now examples showing how to apply differentials when making approximations of the values of trigonometric functions.

Example 1. Find an approximate value of $\sin 31°$.

Solution. Let $x = 30° = \frac{\pi}{6}$, $dx = 1° = \frac{\pi}{180}$, $\sin 30° = \frac{1}{2}$,

$$\Delta\sin x\Big|_{x=\frac{\pi}{6}} \cong d\sin x\Big|_{x=\frac{\pi}{6}} = \cos\frac{\pi}{6}\cdot dx = \frac{\sqrt{3}}{2}\cdot\frac{\pi}{180},$$

$$\sin 31° \cong 0.515115.$$

Actually, with five decimal digits, $\sin 31° \cong 0.51504$.

Example 2. Find an approximate value of tan 44°.

Solution. Let $x = 45° = \frac{\pi}{4}$, $dx = -1° = -\frac{\pi}{180}$. Since tan $45° = 1$, it follows that

$$\Delta \tan x\Big|_{x=\frac{\pi}{4}} \cong d \tan x\Big|_{x=\frac{\pi}{4}} = \frac{1}{\cos^2 \frac{\pi}{4}}\, dx$$

$$= -\frac{\pi}{90} \cong -0.0349,$$

$$\tan 44° \cong 1 - 0.0349 = 0.9651.$$

Actually, with four decimal digits, tan 44° = 0.9657.

5.17. Application of Derivatives to the Investigation of Functions Which Can be Expressed in Terms of Trigonometric Functions

It is sufficient to investigate a periodic function with the least period T on a half-open interval of length T, say, $[0, T)$ or $\left(-\frac{1}{2}T, \frac{1}{2}T\right]$.

Example 1. Construct the graph of the function $y = \sin x + \cos x$.

Solution. The function y is continuous and has a period 2π. We investigate it on the interval $[0, 2\pi)$. At the endpoints of this interval the values of this function are equal to 1. The derivative is $y' = \cos x - \sin x = \cos x\ (1 - \tan x)$. It vanishes for $\tan x = 1$ (and, as can be seen from the nontransformed expression of the derivative, does not vanish where $\cos x = 0$), i.e. on the interval $[0, 2\pi)$ at the points $x = \pi/4$ and $x = 5\pi/4$. If x increases and passes through $\pi/4$ or through $5\pi/4$, then $\tan x$ passes from the values smaller than 1 to the values larger than 1 so that the factor $1 - \tan x$ changes sign from plus to minus. Furthermore, $\cos \frac{\pi}{4} > 0$ and $\cos \frac{5\pi}{4} < 0$. Therefore, when x passes through $\pi/4$, the derivative y' changes sign from plus to minus so that the function y stops increasing and begins to decrease, and has a maximum for $x = \pi/4$. When x passes through $5\pi/4$, y' changes sign from minus to plus,

the function y stops decreasing and begins to increase, and has a minimum for $x = 5\pi/4$. The value at the point of maximum is $\sin\frac{\pi}{4}+\cos\frac{\pi}{4}=\frac{\sqrt{2}}{2}+\frac{\sqrt{2}}{2}=\sqrt{2}$ and the value at the point of minimum is $-\sqrt{2}$. The graph of the function $y = \sin x + \cos x$ on the interval $[0, 2\pi]$ is shown

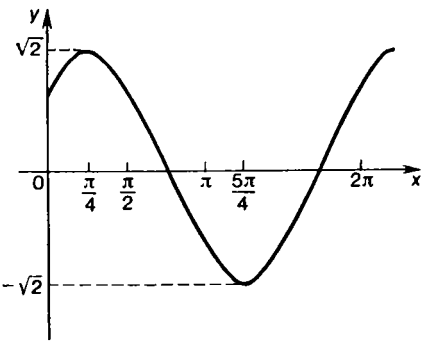


Fig. 5.26

in Fig. 5.26. The graph cuts the abscissa axis at the points $x = 3\pi/4$ and $x = 7\pi/4$.

We could have investigated this function (and we have even done so in a more general form above) by elementary means, without resort to differential calculus. The matter is that

$$\begin{aligned} y &= \sin x + \cos x \\ &= \sqrt{2}\left(\sin x \cos\frac{\pi}{4}+\cos x \sin\frac{\pi}{4}\right) \\ &= \sqrt{2}\sin\left(x+\frac{\pi}{4}\right), \end{aligned}$$

i.e. the graph of this function is a sine curve with period 2π, amplitude $\sqrt{2}$ and with a shift to the left by $\frac{\pi}{4}$.

Example 2. Construct the graph of the function $y = \tan x - \tan 2x$.

Solution. The period of the function is π. It is sufficient to investigate it on the interval $(-\pi/2, \pi/2)$ (the endpoints are not included since the function is not defined at them). The points $x = \pi/4$ and $x = -\pi/4$ are points of discontinuity, at these points $\tan 2x$ suffers discontinuities.

The following table from which we can define fragments of the graph (Fig. 5.27) shows the behaviour of the function

x	$\tan x$	$\tan 2x$	y
$x \to -\frac{\pi}{2}$ from the right	$\to -\infty$	$\to 0$	$\to -\infty$
$x \to -\frac{\pi}{4}$ from the left	$\to -1$	$\to +\infty$	$\to -\infty$
$x \to -\frac{\pi}{4}$ from the right	$\to -1$	$\to -\infty$	$\to +\infty$
$x \to \frac{\pi}{4}$ from the left	$\to 1$	$\to +\infty$	$\to -\infty$
$x \to \frac{\pi}{4}$ from the right	$\to 1$	$\to -\infty$	$\to +\infty$
$x \to \frac{\pi}{2}$ from the left	$\to +\infty$	$\to 0$	$\to +\infty$

$y = \tan x - \tan 2x$ in the vicinity of the endpoints of the interwal $(-\pi/2, \pi/2)$ and in the neighbourhood of the points of discontinuity $-\pi/4$ and $\pi/4$.

From the investigation we have carried out we infer that the function y has at least one maximum on the interval $(-\pi/2, -\pi/4)$ and at least one minimum on the interval $(\pi/4, \pi/2)$.

We introduce a derivative

$$y' = \frac{1}{\cos^2 x} - \frac{2}{\cos^2 2x} = \frac{\cos^2 2x - 2\cos^2 x}{\cos^2 x \cos^2 2x}.$$

The condition $y' = 0$ is equivalent to the equation $\cos^2 2x - 2\cos^2 x = 0$ from which we find that

$$\cos 2x = \pm\sqrt{2}\cos x;\ 2\cos^2 x - 1 = \pm\sqrt{2}\cos x,$$

$$\cos x = \frac{\pm\sqrt{2} \pm \sqrt{2+8}}{4} = \frac{\pm\sqrt{2} \pm \sqrt{10}}{4}.$$

On the interval $(-\pi/2, \pi/2)$ $\cos x > 0$ and therefore we must retain only the values $\cos x = \frac{\sqrt{10} \pm \sqrt{2}}{4}$. The plus sign in this relation can also be dropped since

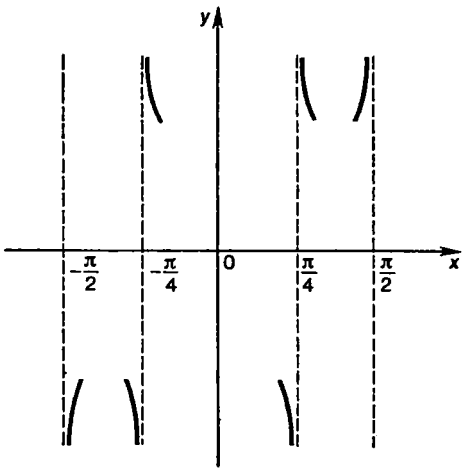


Fig. 5.27

$\frac{\sqrt{10}+\sqrt{2}}{4} > 1$ and the equality $\cos x = \frac{\sqrt{10}+\sqrt{2}}{4}$ is impossible. Thus $\cos x = \frac{\sqrt{10}-\sqrt{2}}{4}$ and $x = \arccos \frac{\sqrt{10}-\sqrt{2}}{4}$, $x = -\arccos \frac{\sqrt{10}-\sqrt{2}}{4}$. It follows from the arguments we have presented that at the point $x = -\arccos \frac{\sqrt{10}-\sqrt{2}}{4}$ the function has a maximum and at the point $x = \arccos \frac{\sqrt{10}-\sqrt{2}}{4}$ it has a minimum. It is easy to find that the value at the point of minimum is equal to $\frac{1}{4}\sqrt{1+\sqrt{5}}\,(\sqrt{10}+3\sqrt{2}) \cong 3.33$. The value at the point of maximum is equal to $-\frac{1}{4}\sqrt{1+\sqrt{5}}\,(\sqrt{10}+3\sqrt{2}) \cong -3.33$.

On the interval $(-\pi/2, \pi/2)$ the graph of the function is shown in Fig. 5.28 and then the graph is periodically re-

peated. The middle descending portion of the graph cuts the abscissa axis at an angle of $-45°$ when $x = 0$.

Example 3. Construct the graph of the function $y = \sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x$.

Solution. The period of the function y is 2π. We investigate its behaviour on the interval $[-\pi, \pi]$. The function is

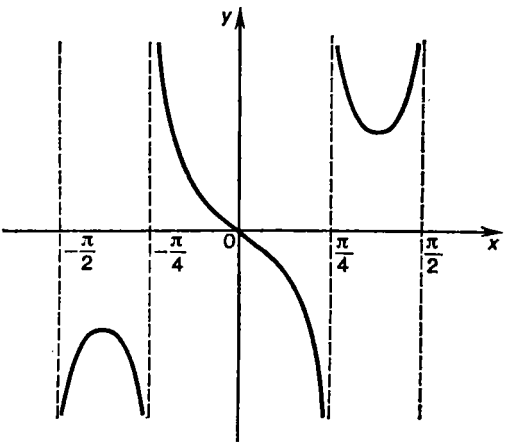


Fig. 5.28

continuous everywhere. For $x = 0$ and at the endpoints $-\pi$ and π of the interval y vanishes. Its derivative is $y' = \cos x + \cos 3x + \cos 5x$. For $x = 0$ the value of the derivative is 3, and so the graph of the function rises very steeply (like the straight line $y = 3x$). For $x = -\pi$ and $x = \pi$ the value of the derivative is -3, and the graph is as steep when x recedes from $-\pi$ as when it approaches π. We transform the derivative y' so as to make it easy to find its roots. With this aim in view, we multiply it by $\sin x$. We obtain

$$(\sin x)\, y' = \sin x \cos x + \sin x \cos 3x + \sin x \cos 5x$$

$$= \frac{1}{2}\sin 2x + \frac{1}{2}(\sin 4x - \sin 2x)$$

$$+ \frac{1}{2}(\sin 6x - \sin 4x) = \sin 6x,$$

$$y' = \frac{\sin 6x}{2 \sin x}\,.$$

On the interval $(-\pi, \pi)$ its roots are evidently

$$-\frac{5}{6}\pi,\ -\frac{2}{3}\pi,\ -\frac{1}{2}\pi,\ -\frac{1}{3}\pi,\ -\frac{1}{6}\pi,\ \frac{1}{6}\pi,\ \frac{1}{3}\pi,$$
$$\frac{1}{2}\pi,\ \frac{2}{3}\pi,\ \frac{5}{6}\pi.$$

It is clear from the form of the derivative before the transformation that the number 0 is not its root. When x increases and passes through any root y', the function $\sin 6x$ and, together with it y', changes sign. Consequently every time y stops increasing and begins to decrease and then again to increase. Therefore all the roots y' are points of maximum and minimum for y. On the interval $[-\pi/6, \pi/6]$ the function y increases (it increases in the neighbourhood of $x = 0$), on the adjacent intervals to the right and to the left it decreases and so on. Therefore $\pi/6$ is a point of maximum, $\pi/3$ is a point of minimum, and so on, $-\pi/6$ is a point of minimum, $-\pi/3$ is a point of maximum and so on. The values at these points are given in the following table.

x	$\frac{\pi}{6} \cong 0.53$	$\frac{\pi}{3} \cong 1.05$	$\frac{\pi}{2} \cong 1.57$	$\frac{2\pi}{3} \cong 2.09$	$\frac{5\pi}{6} \cong 2.62$
y	$\frac{14}{15} \cong 0.93$	$\frac{2\sqrt{3}}{5} \cong 0.69$	$\frac{13}{15} \cong 0.87$	$\frac{2\sqrt{3}}{5} \cong 0.69$	$\frac{14}{15} \cong 0.93$

For the negative values of x the respective values of y differ by signs from those given in the table since the function y is odd. Thus, on $[-\pi, \pi]$ the graph has the form shown in Fig. 5.29.

◀ **Remark.** If we cosidered functions

$$y = \sin x + \frac{1}{3}\sin x + \frac{1}{5}\sin 5x + \frac{1}{7}\sin 7x,$$

$$y = \sin x + \frac{1}{3}\sin 3x + \frac{1}{5}\sin 5x + \frac{1}{7}\sin 7x + \frac{1}{9}\sin 9x$$

and so on, the number of maxima and minima would increase and the values of the maxima and minima within the interval $(0, \pi)$ would differ less and less from one another and would approach the number $\pi/4$. This means that on

this interval the middle part of the graph would approach the straight line $y = \pi/4 \cong 0.785$.

Thus the formula

$$\lim_{n\to\infty}\left(\sin x + \frac{1}{3}\sin 3x + \ldots + \frac{1}{2n-1}\sin(2n-1)x\right) = \frac{\pi}{4}$$

proves to hold true for all x *within* the interval $(0, \pi)$. The same limit for x from the interval $(-\pi, 0)$ is equal to $-\pi/4$.

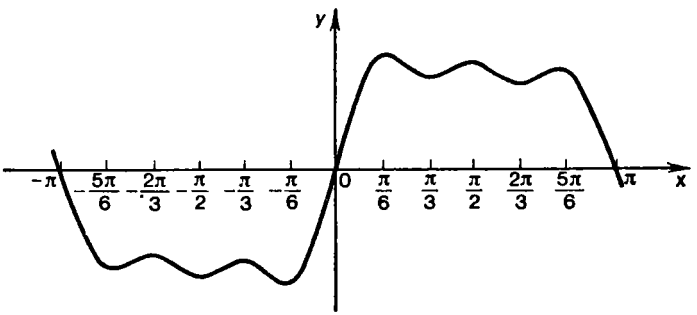


Fig. 5.29

We can thus define a function equal to $\pi/4$ for $0 < x < \pi$ and equal to $-\pi/4$ for $-\pi < x < 0$ by a single formula for the whole interval $(-\pi, \pi)$.

Example 4. Find the number of roots of the equation $3\sin 3x + 4\sin 4x + 5\sin 5x = 0$ on the interval $(-\pi, \pi]$.

Solution. The period of the left-hand side is 2π. The numbers 0 and $\pm\pi$ are roots of the equation, the last two differing by a period. We transform the left-hand side as follows:

$$\sin 3x = \sin(4x - x) = \sin 4x\cos x - \cos 4x\sin x,$$

$$\sin 5x = \sin(4x + x) = \sin 4x\cos x + \cos 4x\sin x,$$

$$3\sin 3x + 4\sin 4x + 5\sin 5x = (8\cos x + 4)\sin 4x + 2\sin x\cos 4x,$$

$$4(2\cos x + 1)\sin 4x + 2\sin x\cos 4x = 0.$$

The roots of the functions $2\cos x + 1$ and $\cos 4x$ are evidently not the roots of the equation being investigated. Therefore, dividing it by $4(2\cos x + 1)\cos 4x$, we arrive at an

equivalent equation

$$F(x) = \tan 4x + \frac{\sin x}{2(2\cos x + 1)} = 0.$$

The function $F(x)$ has discontinuities at the points (in an increasing order)

$$-\frac{7}{8}\pi,\ -\frac{2}{3}\pi,\ -\frac{5}{8}\pi,\ -\frac{3}{8}\pi,\ -\frac{1}{8}\pi,\ \frac{1}{8}\pi,\ \frac{3}{8}\pi,$$
$$\frac{5}{8}\pi,\ \frac{2}{3}\pi,\ \frac{7}{8}\pi.$$

They divide the interval $(-\pi, \pi)$ into eleven parts, two "extremal" parts, $\left(-\pi, -\frac{7}{8}\pi\right)$ and $\left(\frac{7}{8}\pi, \pi\right)$, and nine "middle" parts.

Let us find $F'(x)$. It is equal to

$$\frac{4}{\cos^2 4x} + \frac{2+\cos x}{2(2\cos x+1)^2}.$$

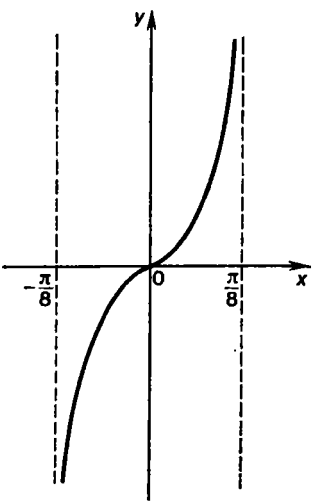


Fig. 5.30

At all points, except for the points of discontinuity, both terms are positive. Therefore, everywhere, except for the points of discontinuity, the function $F(x)$ increases. As x approaches any point of discontinuity, the function $F(x)$ tends to $+\infty$ or $-\infty$ according as x approaches the point of discontinuity from the left or from the right. In each of the middle intervals the graph of the function $F(x)$ has the form shown in Fig. 5.30 and cuts the abscissa axis at one and only one point, so that each of the middle intervals has exactly one root of $F(x)$. In the extreme left interval $[-\pi, -\frac{7}{8}\pi]$ the function $F(x)$ increases from 0 to $+\infty$ and the extreme right interval $(\frac{7}{8}\pi, \pi]$ it increases from $-\infty$ to 0. Thus the numbers $-\pi$ and π are also roots of the equation. They differ by a period.

Thus the equation in question has ten roots on the interval $(-\pi, \pi]$.

It is easy to obtain a similar result for a more general equation

$$a \sin (n-1)x + b \sin nx + c \sin (n+1)x = 0$$

when the conditions $0 < a < c$ and $|b| < a + c$ are satisfied. ▶

5.18. Derivatives and Differentials of Inverse Trigonometric Functions

Let us find the derivatives of the arc sine and arc cosine functions using the rule of differentiation of a function of a function.

Let $y = \arcsin x$. This means that $-\pi/2 \leqslant y \leqslant \pi/2$ and $\sin y = x$. From the last relation we find that $\cos y\, dy = dx$, whence $dy = \frac{1}{\cos y}$, $dy = \frac{1}{\sqrt{1-\sin^2 y}}\, dx$. We put the plus sign before the square root because $\cos y > 0$ for $-\pi/2 < y < \pi/2$. The final result is $dy = \frac{1}{\sqrt{1-x^2}}\, dx$ and

$$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}.$$

This formula holds true for $-1 < x < 1$. We must exclude the endpoints -1 and $+1$ of the interval since $y = -\pi/2$, $y = \pi/2$ and $\cos y = 0$ for them. Note that as $x \to 1$ or $x \to -1$, the value of the derivative $(\arcsin x)'$ tends to infinity. This means that a tangent to the graph of the function $\arcsin x$ is vertical at the endpoints -1 and $+1$ of its domain of definition.

In the same way we can calculate the derivative of the arc tangent. Let $y = \arctan x$. Then $\tan y = x$, $\frac{1}{\cos^2 y}\, dy = dx$. But $\frac{1}{\cos^2 y} = 1 + \tan^2 y = 1 + x^2$, and therefore $dy = d(\arctan x) = \frac{dx}{1+x^2}$ and

$$(\arctan x)' = \frac{1}{1+x^2}.$$

Here we could even do without using the fact that $-\pi/2 < \arctan x < \pi/2$. This is due to the fact that all solutions of the equation $\tan y = x$ differ from $\arctan x$ only by the constant terms $k\pi$ and so their derivatives must coincide.

Example 1. Find the differential and the derivative of the function $y = \arcsin \frac{1}{x}$.

Solution. This function is defined for $1 \leqslant x < \infty$ and for $-\infty < x \leqslant -1$. We have

$$dy = \frac{1}{\sqrt{1-\left(\frac{1}{x}\right)^2}}\, d\left(\frac{1}{x}\right) = \sqrt{\frac{x^2}{x^2-1}}\left(-\frac{dx}{x^2}\right)$$

$$= -\frac{|x|}{\sqrt{x^2-1}} \cdot \frac{dx}{|x|^2} = -\frac{1}{|x|\sqrt{x^2-1}}\, dx,$$

$$y' = -\frac{1}{|x|\sqrt{x^2-1}},$$

i.e. $y' = -\frac{1}{x\sqrt{x^2-1}}$ for $x > 1$ and $y' = \frac{1}{x\sqrt{x^2-1}}$ for $x < -1$.

Example 2. Find the differential and the derivative of the function $y = \arctan \sqrt{x}$.

Solution. The function is defined for $x \geqslant 0$. We have

$$dy = \frac{1}{1+(\sqrt{x})^2} \cdot d\sqrt{x} = \frac{1}{1+x} \cdot \frac{1}{2\sqrt{x}}\, dx,$$

$$y' = \frac{1}{2(1+x)\sqrt{x}}.$$

This is true for $x > 0$, and $y' \to \infty$ as $x \to 0$.

Example 3. Find $y'(x)$ for the function $y = \arctan \frac{1+x}{1-x}$.

Solution. The function is defined everywhere, except for $x = 1$, and $y \to \frac{\pi}{2}$ as $x \to 1$ from the left and $y \to -\frac{\pi}{2}$ as $x \to 1$ from the right. We have

$$dy = \frac{1}{1+\left(\frac{1+x}{1-x}\right)^2}\, d\,\frac{1+x}{1-x}$$

$$= \frac{(1-x)^2}{1-2x+x^2+1+2x+x^2} \cdot \frac{(1-x)+(1+x)}{(1-x)^2}\, dx = \frac{1}{1+x^2}\, dx,$$

$$y' = \frac{1}{1+x^2}.$$

It turned out that the derivative of the function $y = \arctan \frac{1+x}{1-x}$ is equal to the derivative of $z = \arctan x$. It is easy to realize why this is so:

$$\tan\left(z+\frac{\pi}{4}\right)=\frac{\tan z+\tan\frac{\pi}{4}}{1-\tan z\tan\frac{\pi}{4}}=\frac{1+x}{1-x}\,.$$

Taking into account the fact that the values of the arctangent lie in the interval $(-\pi/2, \pi/2)$, we obtain

$$z+\frac{\pi}{4}=\arctan x+\frac{\pi}{4}=\arctan\frac{1+x}{1-x}\quad\text{for } x<1$$

and

$$\arctan x+\frac{\pi}{4}=\arctan\frac{1+x}{1-x}+\pi\quad\text{for } x>1.$$

In both cases the functions $\arctan \frac{1+x}{1-x}$ and $\arctan x$ differ by constant terms.

EXERCISES TO CHAPTER 5

1. Calculate the sine, the cosine, the tangent and the cotangent of the angles of 15°, 75° and 105°.

2. Calculate the value of each of the following expressions:

1. $\sin 12^\circ \cos 33^\circ + \cos 12^\circ \sin 33^\circ$.
2. $\cos 100^\circ \sin 10^\circ - \sin 100^\circ \cos 10^\circ$.
3. $\cos\frac{\pi}{12}\cos\frac{\pi}{4}-\sin\frac{\pi}{12}\sin\frac{\pi}{4}$.
4. $\sin\frac{\pi}{5}\sin\frac{\pi}{20}+\cos\frac{\pi}{5}\cos\frac{\pi}{20}$.
5. $\sin\left(\frac{\pi}{3}-\alpha\right)\cos\left(\frac{\pi}{6}+\alpha\right)+\sin\left(\frac{\pi}{6}+\alpha\right)\cos\left(\frac{\pi}{3}-\alpha\right)$.
6. $\sin\left(\frac{\pi}{4}+\alpha\right)\sin\left(\frac{\pi}{4}-\alpha\right)-\cos\left(\frac{\pi}{4}+\alpha\right)\cos\left(\frac{\pi}{4}-\alpha\right)$.

3. Prove that the sine of the sum of positive numbers smaller than $\pi/2$ is smaller than the sum of the sines of those numbers. Give a graphical interpretation of this fact on a unit circle.

4. About which of the points, $\alpha = 0^\circ$ or $\alpha = 45^\circ$, will the increment $\Delta\alpha = 1^\circ$ yield a larger increment of the function $y = \sin\alpha$? Use a unit circle to find the answer and then verify your reasoning using the result of the preceding problem.

5. Given that $\sin \alpha = a$, $\sin \beta = b$, α and β are angles of a triangle, calculate the sine of the third angle of the triangle if:
(a) α and β are acute angles,
(b) one of the angles is abtuse.

6. Given that $\cos \alpha = m$, $\cos \beta = n$, α and β are angles of a triangle, calculate the cosine of the third angle of the triangle.

7. Reduce the following expressions to the form $A \sin (at + \varphi)$ or $A \cos (at + \varphi)$, where $A > 0$ and $0 \leqslant \varphi < 2\pi$. Find the greatest value of each of the expressions:

1. $\frac{\sqrt{2}}{2} \sin t + \frac{\sqrt{2}}{2} \cos t$. 2. $\frac{\sqrt{2}}{2} \cos 2t - \frac{\sqrt{2}}{2} \sin 2t$.

3. $\frac{1}{2} \sin \frac{t}{2} - \frac{\sqrt{3}}{2} \cos \frac{t}{2}$. 4. $\frac{\sqrt{3}}{2} \sin t - \frac{1}{2} \cos t$.

5. $-3 \sin t - 4 \cos t$. 6. $-4 \sin t - 3 \cos t$.

7. $\sin \left(3t - \frac{\pi}{4}\right) + \cos \left(3t - \frac{\pi}{4}\right)$.

8. $\cos \left(\frac{1}{3} t + \frac{\pi}{3}\right) - \sin \left(\frac{1}{3} t + \frac{\pi}{3}\right)$.

8. Find the values of the variable t for which the expressions given in Problem 7 assume:
(a) the greatest value,
(b) the least value,
(c) a value equal to zero.

9. Simplify the following expressions:

1. $\frac{\sin (\alpha + \beta) - \sin \beta \cos \alpha}{\sin (\alpha - \beta) + \sin \beta \cos \alpha}$.

2. $\frac{\cos (\alpha + \beta) + \sin \alpha \sin \beta}{\cos (\alpha - \beta) - \sin \alpha \sin \beta}$.

3. $\cos \left(\alpha + \frac{\pi}{6}\right) + \cos \left(\alpha - \frac{\pi}{6}\right)$.

4. $\cos \left(\alpha + \frac{\pi}{3}\right) - \cos \left(\alpha - \frac{\pi}{3}\right)$.

10. Prove the following identities:

1. $\sin (\alpha - \beta) \sin (\alpha + \beta) = \cos^2 \beta - \cos^2 \alpha$.
2. $\cos (\alpha + \beta) \cos (\alpha - \beta) = \cos^2 \beta - \sin^2 \alpha$.
3. $\tan \alpha + \cot \beta = \frac{\cos (\beta - \alpha)}{\sin \beta \cos \alpha}$.
4. $\cot \alpha - \tan \beta = \frac{\cos (\alpha + \beta)}{\sin \alpha \cos \beta}$.

11. Calculate the following expressions using the formula for the tangent of the sum or the difference:

$\tan 15°$, $\tan 75°$, $\tan 105°$, $\cot \frac{\pi}{12}$, $\cot \frac{5}{12}\pi$, $\cot \frac{7}{12}\pi$.

12. Calculate $\tan\left(\frac{\pi}{4}+\alpha\right)$ if $\sin\alpha=\frac{3}{5}$ and $\frac{\pi}{2}<\alpha<\pi$.

13. Calculate $\tan\left(\alpha-\frac{\pi}{3}\right)$ if $\cos\alpha=-0.8$ and $\pi<\alpha<\frac{3}{2}\pi$.

14. Calculate:

1. $\dfrac{\tan 17°+\tan 28°}{1-\tan 17°\tan 28°}$. 2. $\dfrac{\tan 70°-\tan 10°}{1+\tan 70°\tan 10°}$.

3. $\dfrac{\tan\frac{\pi}{42}+\tan\frac{\pi}{7}}{1-\tan\frac{\pi}{7}\tan\frac{\pi}{42}}$. 4. $\dfrac{1+\tan\frac{\pi}{12}}{1-\tan\frac{\pi}{12}}$.

15. Prove the following identities:

1. $\dfrac{\tan^2\alpha-\tan^2\beta}{1-\tan^2\alpha\tan^2\beta}=\tan(\alpha+\beta)\tan(\alpha-\beta)$.

2. $\dfrac{\tan\alpha+\tan\beta}{\tan(\alpha+\beta)}+\dfrac{\tan\alpha-\tan\beta}{\tan(\alpha-\beta)}=2$.

16. Given that $\tan\alpha=\frac{1}{2}$, $\tan\beta=\frac{1}{3}$, $\tan\gamma=\frac{1}{7}$, α, β, γ are acute angles, prove that $\alpha=\beta+\gamma$.

17. Given that $\tan\alpha=\frac{1}{12}$, $\tan\beta=\frac{2}{5}$, $\tan\gamma=\frac{1}{3}$, α, β, γ are acute angles, prove that $\alpha+\beta+\gamma=45°$.

18. Calculate $\sin x$, $\cos x$, $\tan x$, $\cot x$ if:

1. $\sin\frac{x}{2}=-0.6$, $-\frac{\pi}{2}<\frac{x}{2}<0$.

2. $\cos\frac{x}{2}=-0.8$, $-\pi<\frac{x}{2}<-\pi/2$.

3. $\cos\frac{x}{2}=-\frac{12}{13}$, $\frac{\pi}{2}<\frac{x}{2}<\pi$.

4. $\sin\frac{x}{2}=\frac{5}{13}$, $0<\frac{x}{2}<\frac{\pi}{2}$.

19. Solve the following equations:

1. $\sin x-\sin\frac{x}{2}=0$. 2. $\cos\frac{z}{2}+\sin z=0$.

3. $1-\cos 2t+\sin t=0$. 4. $\cos x+1+\cos 2x=0$.

5. $\tan\varphi+\tan 2\varphi=0$. 6. $\cot\psi-\cot 2\psi=0$.

7. $\sin\left(\frac{\pi}{3}+\frac{\alpha}{2}\right)\sin\left(\frac{\pi}{6}-\frac{\alpha}{2}\right)=-\frac{1}{2}$.

8. $\cos\left(\frac{\pi}{4}-\frac{\beta}{2}\right)\cos\left(\frac{\pi}{4}+\frac{\beta}{2}\right)=\frac{1}{2}$.

9. $\cos^2\left(\frac{2}{3}\pi+\varphi\right)-\cos^2\left(\frac{\pi}{6}+\varphi\right)=1$.

10. $\cos^2\left(\frac{5}{6}\pi-\alpha\right)-\cos^2\left(\frac{\pi}{3}-\alpha\right)=-1$.

20. Calculate $\sin x$, $\cos x$, $\tan x$, $\cot x$ if:

1. $\tan\frac{x}{2}=\frac{3}{4}$, $0<\frac{x}{2}<\frac{\pi}{2}$.

2. $\tan\frac{x}{2}=-\frac{1}{4}$, $-\frac{\pi}{2}<\frac{x}{2}<0$.

3. $\tan\frac{x}{2}=-7$, $\frac{\pi}{2}<\frac{x}{2}<\pi$.

4. $\tan\frac{x}{2}=2$, $\pi<\frac{x}{2}<\frac{3}{2}\pi$.

21. Find the greatest value of each of the following expressions and the value of the variable for which it is attained:

1. $\sin x\cdot\cos x$. 2. $(\cos x+\sin x)(\cos x-\sin x)$.
3. $(\sin x+\cos x)^2$. 4. $(\sin x-\cos x)^2$.

22. Proceeding from the graphs of the functions $y=\sin x$ and $y=\cos x$ and the formula for the sine of the doubled argument, construct the graph of the function $y=\sin 2x$.

23. Prove the following indentities:

1. $\cos^4\alpha-\sin^4\alpha=\cos 2\alpha$.

2. $\sin\frac{x}{4}\cos\frac{x}{4}\cos\frac{x}{2}\cos x\cos 2x=\frac{1}{16}\sin 4x$.

3. $\frac{1+\sin 2\alpha+\cos 2\alpha}{1+\sin 2\alpha+\cos 2\alpha}=\cot\alpha$. 4. $\frac{\sin 2\alpha}{1+\cos 2\alpha}=\tan\alpha$.

5. $\cos^3\alpha\sin\alpha-\sin^3\alpha\cos\alpha=\frac{1}{4}\sin 4\alpha$.

6. $\frac{1-\cos z+\cos 2z}{\sin 2z-\sin z}=\cot z$.

24. Calculate $\sin x$, $\cos x$, $\tan x$ and $\cot x$ if:

1. $\sin 2x=0.6$, $\frac{\pi}{2}<2x<\pi$. 2. $\cos 2x=-0.8$, $\pi<2x<2\pi$.

3. $\tan 2x=-4$, $-\frac{\pi}{2}<2x<0$.

4. $\cot 2x=\frac{1}{2}$, $-\pi<2x<-\frac{\pi}{2}$.

25. Construct the graphs of the relations:

1. $\sin(x+y)=\sin x+\sin y$. 2. $\sin(x-y)=\sin x-\sin y$.

26. Give a linear notation for each of the following expressions in terms of trigonometric functions of multiple arguments:

1. $\cos^4\alpha$. 2. $\sin^4\alpha$. 3. $(\sin x+\cos x)^4$.
4. $(\sin x-\cos x)^4$.

27. Prove the following equalities:

1. $\sin 35^\circ+\sin 25^\circ=\cos 5^\circ$. 2. $\cos 10^\circ-\cos 50^\circ=\sin 20^\circ$.
3. $\cos 40^\circ+\cos 80^\circ=\cos 20^\circ$.
4. $\sin 50^\circ-\sin 70^\circ=-\sin 10^\circ$.
5. $\cos 45^\circ-\sin 75^\circ=-\sin 15^\circ$.
6. $\cos 20^\circ-\sin 10^\circ=\sin 50^\circ$.

28. Construct the graphs of the following relations:

1. $\cos(x+y)=\cos x+\cos y-1$.
2. $\cos(x-y)=\cos x-\cos y+1$.

29. Solve the following equations:

1. $\sin x+\sin 3x=0$. 2. $\cos x+\cos 3x=0$.
3. $\cos\frac{x}{2}-\cos\frac{3}{2}x=0$. 4. $\sin\frac{x}{2}-\sin\frac{3}{2}x=0$.
5. $\sin 2x+\cos 4x=0$. 6. $\cos 4x-\sin 2x=0$.

30. Prove the following identities:

1. $\sin^2(\alpha+\beta)-\sin^2(\alpha-\beta)=\sin 2\alpha\sin 2\beta$.
2. $\cos^2(\alpha-\beta)-\cos^2(\alpha+\beta)=\sin 2\alpha\sin 2\beta$.
3. $\dfrac{\sin\alpha+\sin 3\alpha+\sin 5\alpha+\sin 7\alpha}{\cos\alpha+\cos 3\alpha+\cos 5\alpha+\cos 7\alpha}=\tan 4\alpha$.
4. $\dfrac{\sin\alpha+\sin 2\alpha+\sin 3\alpha+\sin 4\alpha}{\cos\alpha+\cos 2\alpha+\cos 3\alpha+\cos 4\alpha}=\tan\frac{5}{2}\alpha$.

31. Prove the following identities under the condition that α, β and γ are angles of a triangle:

1. $\sin\alpha+\sin\beta+\sin\gamma=4\cos\frac{\alpha}{2}\cos\frac{\beta}{2}\cos\frac{\gamma}{2}$.
2. $\sin\alpha+\sin\beta-\sin\gamma=4\sin\frac{\alpha}{2}\sin\frac{\beta}{2}\cos\frac{\gamma}{2}$.
3. $\tan\alpha+\tan\beta+\tan\gamma=\tan\alpha\tan\beta\tan\gamma$.
4. $\cot\frac{\alpha}{2}+\cot\frac{\beta}{2}+\cot\frac{\gamma}{2}=\cot\frac{\alpha}{2}\cot\frac{\beta}{2}\cot\frac{\gamma}{2}$.

32. Solve the following equations:

1. $\sin x\sin 7x=\sin 3x\sin 5x$.
2. $\cos x\sin 7x=\cos 3x\sin 5x$.

33. Construct the graphs of the following functions:

1. $y = \sin x + \cos x$. 2. $y = \cos x - \sin x$.
3. $y = -\sqrt{3}\sin t - \cos t$. 4. $y = \sin z + \sqrt{3}\cos z$.
5. $y = 2\cos x - 3\sin x$. 6. $y = 3\sin t - 4\cos t$.

34. Find an approximate value for the root closest to zero in each of the following equations:

1. $5\sin x + 12\cos x = 13$. 2. $7\sin x + 24\cos x = -25$.
3. $\cos 2x - 3\sin 2x = -\sqrt{10}$.
4. $3\cos 2x - \sin 2x = \sqrt{10}$.

35. A point M moves along the circumference of a circle, with centre at the origin, in the positive direction with a constant linear velocity of 4 m/s. The radius of the circle $R = 2$ m. Write the law of variation of the coordinates of the point M as a function of time if its ordinate is known to be equal to 2 m for $t = \frac{1}{2}$ s. Find the period and the frequency of the harmonic oscillations which the coordinates of the point undergo, and their initial phase. Explain the physical meaning of the coefficient of t in the notation of the law of variation of the coordinates.

36. Find the sums of the harmonic oscillations and determine their period, frequency, initial phase and amplitude:

1. $\sin(2t + 1.25\pi) + \cos(2t + 1.75\pi)$.
2. $\cos\left(\frac{1}{2}t + 0.25\pi\right) - \sin\left(\frac{1}{2}t + 0.75\pi\right)$.
3. $2\cos(100\pi t) - 4\sin\left(100\pi t + \frac{2}{3}\pi\right)$.
4. $10\sin(120\pi t) + 8\cos\left(120\pi t + \frac{4}{3}t\right)$.

37. Reduce the equation for the harmonic oscillations to the form $A\cos(\omega t + \varphi)$, where $A > 0$, $\omega > 0$, $0 \leqslant \varphi < 2\pi$:

1. $y = -3\sin(2t - 1.7\pi)$.
2. $y = -4\cos(3.2\pi - 0.5t)$.
3. $y = -6\cos(3t + 1.8\pi)$. 4. $y = -2\sin(5.5\pi - 4t)$.
5. $y = \cos(1.5\pi - 0.3t)\sin(1.5\pi - 0.3t)$.
6. $y = \sin^2(1.5\pi - 1.5t) - \cos^2(1.5\pi - 1.5t)$.

38. Substantiate the following equalities:

1. $\arcsin 1 = \pi/2$. 2. $\arcsin(-1) = \pi/2$.

39. Substantiate the following equalities:

1. $\arcsin(-1/2) = -\pi/6$. 3. $\arcsin(\sqrt{3}/2) = \pi/3$.
2. $\arcsin(\sqrt{2}/2) = \pi/4$. 4. $\arcsin 0 = 0$.

40. On a sheet of millimetre graph paper construct a unit circle taking 4 cm to be a scale unit. Using a protractor, find the approximate value of arcsin a; write the answer in radian and degree measures:

1. arcsin 0.3. 2. arcsin (—0.8). 3. arcsin (—0.45).
4. arcsin 0.25. 5. arcsin 0.75. 6. arcsin (—0.65).

41. In the same system of coordinates construct the graphs of the functions $y = \sin x$ and $y = \arcsin x$ taking 4 cm to be a scale unit. Find arcsin a for the value of a from the preceding problem with the aid of each graph. Compare the answers with the tabular ones.

42. Write the answers using the symbol arcsin a in degree and radian measures. Find the approximate values of the roots on the interval $[0, T]$, where T is the period of the function appearing on the left-hand side of the equation:

1. $\sin x = 0.4$. 2. $\sin x = -0.6$. 3. $\sin 2x = -0.5$.
4. $\sin \frac{1}{2} = 0.5$. 5. $\sin (x - 1) = 0.7$.
6. $\sin (x + 2) = -0.3$.
7. $\sin \left(\frac{x}{3} + \frac{\pi}{3}\right) = -0.25$.
8. $\sin \left(3x + \frac{\pi}{4}\right) = 0.45$.

43. Fill in the following tables:

a	0	-1	1	$-\frac{1}{2}$	$\frac{1}{2}$	$-\frac{\sqrt{2}}{2}$	$\frac{\sqrt{2}}{2}$	$-\frac{\sqrt{3}}{2}$	$\frac{\sqrt{3}}{2}$
arccos a									

a	0	-1	1	$-\sqrt{3}$	$\sqrt{3}$	$-\frac{\sqrt{3}}{3}$	$\frac{\sqrt{3}}{3}$
arctan a							
arccot a							

44. Carry out an assignment similar to that from Problem 40 but for arccos a.

45. Construct the graphs of the following functions:

1. $y = \arccos x$. 2. $y = \arctan x$. 3. $y = \operatorname{arccot} x$.

46. Carry out an assignment given in Problem 42 first reducing the left-hand sides of the equations to the form $\cos (kx + b)$. Use the symbol arccos a to write the exact answers.

47. Write the answer to each of the following equations using the symbol arctan a in degree and radian measures. Find the approximate values of the roots on the interval $[0, T)$, where T is the period of the left-hand side of the equation:

1. $\tan x = 2.5$. 2. $\tan x = -4$. 3. $\tan \frac{x}{2} = -0.5$.

4. $\tan 2x = 2$. 5. $\tan (2 + x) = \frac{2}{3}$.

6. $\tan (x - 1) = -\frac{3}{2}$. 7. $\tan \left(3x - \frac{\pi}{3}\right) = -1\frac{1}{3}$.

8. $\tan \left(\frac{x}{3} - \frac{\pi}{6}\right) = 0.75$.

48. Carry out the assignment given in Problem 47 first reducing the left-hand sides of the equations to the form cot $(kx + b)$. Use the function arccot x to write the answers.

49. Prove that:

(1) $\operatorname{arccos}(-x) = \pi - \operatorname{arccos} x$. (2) $\operatorname{arccot}(-x) = \pi - \operatorname{arccot} x$.

50. Find x from each of the following equations:

1. $\arcsin (x + 2) = \pi/3$.

2. $\arcsin (1 - 2x) = -\pi/6$.

3. $\arccos \frac{x+1}{2} = \frac{\pi}{2}$. 4. $\arccos \frac{3-x}{3} = 0$.

5. $\arctan (x^2 - 1) = 0$. 6. $\arctan (x^2 + 1) = \pi/4$.

51. Prove:

1. $\arcsin \frac{5}{13} + \arcsin \frac{12}{13} = \frac{\pi}{2}$.

2. $\arccos \sqrt{\frac{2}{3}} - \arccos \frac{\sqrt{6}+1}{2\sqrt{3}} = \frac{\pi}{6}$.

3. $\arctan \frac{2}{3} + \arctan \frac{1}{5} = \frac{\pi}{4}$.

4. $\operatorname{arccot} \frac{1}{9} + \operatorname{arccot} \frac{4}{5} = \frac{3\pi}{4}$.

52. Solve each of the following equations on some interval equal to the period and find representatives of all classes of solutions:

1. $\cos 2x + \cos x = 0$. 2. $\sqrt{2} \sin^2 x + \cos x = 0$.

3. $2 \sin^2 x + 3 \sin x \cos x + \cos^2 x = 0$.

4. $\sqrt{3} \sin x \cos x - \sin^2 x = 0$.

5. $3 \sin^2 x - 4 \sin x \cos x + 5 \cos^2 x = 0$.

6. $8 \sin^2 \frac{x}{2} + 3 \sin x - 4 = 0$. 7. $\cos x - \cos 2x = \sin 3x$.

8. $\cos 2x - \cos 8x + \cos 6x = 1$. 9. $\sin 2x + \cos 2x = -1$.

10. $\sin \frac{x}{2} - \cos \frac{x}{2} = 1$. 11. $\sqrt{3} \sin x - \cos x = 2 \sin 3x$.

12. $3 \cos x + 4 \sin x = 5 \cos 7x$.

13. $\sin \left(x - \frac{\pi}{2}\right) - \sin x = \sin 2 \left(x + \frac{\pi}{4}\right)$.

14. $1 - \tan x = \cos 2x$. 15. $\frac{1}{\tan x} - 1 = \frac{\cos 2x}{1 + \tan x}$.

16. $1 + \cos 4x = \frac{1}{1 - \tan^2 2x}$.

53. Use graphical means to illustrate the following inequalities and write the answers:

1. $\sin x < \frac{1}{2}$. 2. $\cos x > -\frac{1}{2}$.

3. $\sin x \geqslant -\frac{\sqrt{3}}{2}$. 4. $\cos x \leqslant \frac{\sqrt{3}}{2}$.

5. $\tan x \leqslant 1$. 6. $\tan x \geqslant -1$.

7. $-\frac{1}{2} < \sin x < \frac{\sqrt{3}}{2}$. 8. $-\frac{\sqrt{2}}{2} \leqslant \cos x \leqslant \frac{1}{2}$.

54. Solve the following inequalities:

1. $\sin x \cos 2x < 0$. 2. $\cos x \sin 2x > 0$.

3. $3 \sin^2 x - 10 \sin x + 3 \geqslant 0$.

4. $4 \cos^2 x + 17 \cos x + 4 \leqslant 0$.

55. Find the intervals of monotonicity of each of the following functions:

1. $y = x + \sin x$. 2. $y = x - \cos x$.

3. $y = \sin x \tan x$. 4. $y = \cos x \cot x$.

56. Construct the graphs of the following functions (e is the Napier's number):

1. $y = \cos x - \cos^2 x$. 2. $y = \frac{1}{\sin x + \cos x}$.

3. $y = \ln \sin x$. 4. $y = \ln \cos x$. 5. $y = \sqrt{\sin x}$.

6. $y = \sqrt{\cos x}$. 7. $y = e^{\sin x}$. 8. $y = e^{\cos x}$.

57. The point $M(x, y)$ moves along a plane and at each moment of time t, $t \geqslant 0$, its coordinates x and y can be found from the formulas

given below. Depict the path of motion of the point in each case:

1. $x = \cos t$, $y = \sin t$.
2. $x = 0.5 \cos t$, $y = 0.5 \sin t$.
3. $x = 2 \cos t$, $y = 2 \sin t$.
4. $x = 2 \cos t$, $y = \sin t$.
5. $x = \cos t$, $y = 2 \sin t$.

58. Find the greatest value of the function

$$y = \sin^2 x - \sin x + 1$$

on the interval $(\pi, 2\pi)$.

59. Find the least value of the function

$$y = \cos^2 x - \cos x + 1$$

on the interval $(\pi, 2\pi)$.

Chapter 6

ELEMENTS OF INTEGRAL CALCULUS

6.1. Definition of Integration

Integration is the inverse of differentiation, i.e. finding a function from the given derivative or a given differential.

Assume that $f(x)$ is the derivative of the function $F(x)$ at all points of an interval or, what is the same, $f(x)\,dx$ is the differential of the function $F(x)$ in the same interval. Then the function $F(x)$ is called an *antiderivative* for the function $f(x)$ or for the differential $f(x)\,dx$. The process of finding an antiderivative of a function is integration.

Thus the function x^2 is an antiderivative of the function $2x$ (or, what is the same, of the differential $2x\,dx$) throughout the real axis. But it is not the only antiderivative. Other antiderivatives are $x^2 + 2$, $x^2 - 3$, and so on, and, in general, $x^2 + C$, where C is any constant.

It is clear that if $F_0(x)$ is an antiderivative of $f(x)$, then all functions $F_0(x) + C$ for any constant C are antiderivatives of $f(x)$. Indeed,

$$(F_0(x) + C)' = F_0'(x) + C' = f(x) + 0 = f(x).$$

It turns out that all antiderivatives of the function $f(x)$ have the form $F_0(x) + C$, i.e. the following theorem is true.

Theorem. *If $F(x)$ and $F_0(x)$ are two antiderivatives for the same function $f(x)$ on an interval, then they differ by a constant term, i.e. $F(x) = F_0(x) + C$, where C is a constant.*

We shall first formulate and explain the following lemma.

Lemma. *If the function $\varphi(x)$ has a derivative at all points of the interval $[a, b]$ and that derivative is equal to zero throughout the interval, then the function $\varphi(x)$ is a constant.*

It is not easy to present a rigorous proof of this lemma and we shall give it in the next section. For the time being, we shall only give the following explanation. The derivative is the rate of growth of a function. By the hypothesis of the lemma, it is equal to zero at all points of the interval and it is therefore almost obvious that the values of the function do not change when we pass from point to point, otherwise

the rate of growth of the function would somewhere be nonzero.

Now the validity of the theorem follows directly from the lemma. The function $\varphi(x) = F(x) - F_0(x)$ has a derivative $\varphi'(x) = F'(x) - F_0'(x) = f(x) - f(x) = 0$, and is true throughout the interval. By virtue of the lemma, $\varphi(x) = C$ and $F(x) = F_0(x) + C$, and this is what we wished to prove.

An antiderivative of the differential $f(x)\,dx$, written in the general form, is an *indefinite integral* of $f(x)\,dx$ and is designated as $\int f(x)\,dx$. Thus

$$\int f(x)\,dx = F_0(x) + C,$$

where $F_0(x)$ is one of the antiderivatives.

The symbol $\int$ is the *sign of an integral*. This extended and stylized letter S is the first letter of the word "sum". The expression $f(x)\,dx$ appearing under the integral sign is the *element of integration*, and the function $f(x)$ is the *integrand function*.

It is clear from the definition that

$$d\left(\int f(x)\,dx\right) = f(x)\,dx, \quad \int dF(x) = F(x) + C.$$

◀ 6.2. A More Rigorous Proof of the Lemma

We shall now prove the lemma from 6.1 using the mean value theorem or Lagrange's theorem of differential calculus (Lemma 2 in 3.2). Recall the mean value theorem: if $\varphi(x)$ is differentiable at all points of the interval $[a, b]$, then there is a point c in this interval such that $\frac{\varphi(b) - \varphi(a)}{b - a} = \varphi'(c)$.

It is clear that if $\varphi(x)$ is differentiable at all points of the interval $[a, b]$, then this condition is also satisfied for any interval $[x_1, x_2]$ contained in $[a, b]$. Consequently, for any x_1 and x_2 from $[a, b]$, $x_1 < x_2$, there is a point x_3 belonging to the interval $[x_1, x_2]$, which is sure to belong to $[a, b]$, for which $\frac{\varphi(x_2) - \varphi(x_1)}{x_2 - x_1} = \varphi'(x_3)$. By the hypothesis of the lemma, the function $\varphi'(x)$ is zero at all points of the interval $[a, b]$, in particular, $\varphi'(x_3) = 0$. Consequent-

ly, for any x_1 and x_2 from $[a, b]$, $x_1 < x_2$, we have $\frac{\varphi(x_2)-\varphi(x_1)}{x_2-x_1} = 0$, whence it follows that $\varphi(x_2) = \varphi(x_1)$ for any x_1 and x_2 from $[a, b]$.

Thus the values of the function $\varphi(x)$ at all points of the interval $[a, b]$ coincide and, consequently, the function $\varphi(x)$ is a constant. ▶

6.3. Basic Integration Formulas

In Chapter 2 we established rules and formulas for differentiation which make it possible to find derivatives and differentials of functions defined by algebraic expressions. In Chapter 5 we derived formulas for differentiation of trigonometric and inverse trigonometric functions. These rules and formulas yield the following rules and formulas for integration.

1. *The integral of the sum of two functions is equal to th sum of the integrals of the terms.*

Indeed, let $dF_1(x) = f_1(x)\,dx$ and $dF_2(x) = f_2(x)\,dx$. Then

$$d(F_1(x) + F_2(x)) = dF_1(x) + dF_2(x) = f_1(x)\,dx + f_2(x)\,dx$$

and so $F_1(x) + F_2(x)$ is one of the antiderivatives of the functions for $(f_1(x) + f_2(x))\,dx$. This means that

$$\int [f_1(x) + f_2(x)]\,dx = F_1(x) + F_2(x) + C.$$

But $\int f_1(x)\,dx = F_1(x) + C_1$, $\int f_2(x)\,dx = F_2(x) + C_2$ and

$$\int f_1(x)\,dx + \int f_2(x)\,dx = F_1(x) + F_2(x) + (C_1 + C_2)$$
$$= \int (f_1(x)\,dx + f_2(x))\,dx.$$

The sum $C = C_1 + C_2$ of two arbitrary constants is again an arbitrary constant.

Thus

$$\int (f_1(x) + f_2(x))\,dx = \int f_1(x)\,dx + \int f_2(x)\,dx. \qquad (1)$$

2. *The constant factor can be put before the integral sign*, i.e.

$$\int mf(x)\,dx = m\int f(x)\,dx, \tag{2}$$

where m is a constant quantity.

Indeed, if $\int f(x)\,dx = F(x) + C$, then $dm\,F(x) = md\,F(x) = mf(x)\,dx$, so that $mF(x)$ is one of the antiderivatives of the function $mf(x)\,dx$, i.e. $\int mf(x)\,dx = mF(x) + C = m\left(F(x) + \frac{C}{m}\right) = m\int f(x)\,dx$ since $\frac{C}{m}$ is an arbitrary constant.

Let us now consider other differentiation formulas to obtain from them the corresponding integration formulas.

We have $dx^a = ax^{a-1}\,dx$. From this we infer that $\int x^{a-1}\,dx = \frac{1}{a}x^a + C$ and, replacing $a-1$ by a, we get

$$\int x^a\,dx = \frac{1}{a+1}x^{a+1} + C, \tag{3}$$

where $a \neq -1$.

Furthermore, we have $d\ln x = \frac{1}{x}dx$, from which it follows that

$$\int \frac{1}{x}\,dx = \ln x + C. \tag{4}$$

It is interesting to point out that formula (3) loses sense for $a = -1$ and formula (4) compensates for this loss.

From $de^x = e^x dx$ we infer that

$$\int e^x\,dx = e^x + C. \tag{5}$$

From $d\sin x = \cos x\,dx$ it follows that

$$\int \cos x\,dx = \sin x + C. \tag{6}$$

From $d\cos x = -\sin x\,dx$ it follows that

$$\int \sin x\,dx = -\cos x + C. \tag{7}$$

From $d \tan x = \frac{dx}{\cos^2 x}$ it follows that

$$\int \frac{dx}{\cos^2 x} = \tan x + C. \tag{8}$$

From $d \cot x = -\frac{dx}{\sin^2 x}$ it follows that

$$\int \frac{dx}{\sin^2 x} = -\cot x + C. \tag{9}$$

From $d \arcsin x = \frac{dx}{\sqrt{1-x^2}}$ it follows that

$$\int \frac{dx}{\sqrt{1-x^2}} = \arcsin x + C. \tag{10}$$

From $d \arctan x = \frac{dx}{1+x^2}$ it follows that

$$\int \frac{dx}{1+x^2} = \arctan x + C. \tag{11}$$

Formulas (1)-(11) allow us to calculate many simple integrals.

Example 1. $\int (x^2+1)^2\, dx$.

Solution.

$$\int (x^2+1)^2\, dx = \int (x^4+2x^2+1)\, dx = \frac{x^5}{5} + \frac{2x^3}{3} + x + C.$$

Example 2. $\int \frac{x+1}{\sqrt{x}}\, dx$.

Solution.

$$\int \frac{x+1}{\sqrt{x}}\, dx = \int \frac{x}{\sqrt{x}}\, dx + \int \frac{dx}{\sqrt{x}}$$

$$= \int x^{\frac{1}{2}}\, dx + \int x^{-\frac{1}{2}}\, dx$$

$$= \frac{x^{3/2}}{3/2} + \frac{x^{1/2}}{1/2} + C = \frac{2}{3}\sqrt{x^3} + 2\sqrt{x} + C.$$

Example 3. $\int \frac{x^2+x+1}{x}\, dx$.

Solution.

$$\int \frac{x^2+x+1}{x}\, dx = \int \left(x+1+\frac{1}{x}\right) dx = \frac{x^2}{2} + x + \ln x + C.$$

Example 4. $\int \frac{x^2\,dx}{1+x^2}$.

Solution.

$$\int \frac{x^2\,dx}{1+x^2} = \int \frac{(1+x^2)-1}{1+x^2}\,dx = \int \left(1 - \frac{1}{1+x^2}\right) dx$$
$$= x - \arctan\, x + C.$$

Example 5. $\int \tan^2 x\,dx$.

Solution.

$$\int \tan^2 x\,dx = \int \frac{\sin^2 x}{\cos^2 x}\,dx = \int \frac{1-\cos^2 x}{\cos^2 x}\,dx$$
$$= \int \left(\frac{1}{\cos^2 x} - 1\right) dx = \tan x - x + C.$$

6.4. Integration Based on the Use of the Invariance of the Formula for the Differential of a Function of a Function

This technique is based on the invariance of the formula for the differential when the argument is replaced by the function of another argument, i.e. if $dF(x) = f(x)\,dx$, then $dF(\varphi(x)) = f(\varphi(x))\,d\varphi(x)$.

Therefore, if $F(x)$ is an antiderivative for the differential $f(x)\,dx$, then $F(\varphi(x))$ is an antiderivative for the differential $f(\varphi(x))\,d\varphi(x)$.

Let us consider some examples.

Example 1.

$$\int (3x+1)^3\,dx = \frac{1}{3}\int (3x+1)^3\,d(3x+1) = \frac{1}{3}\cdot\frac{(3x+1)^4}{4} + C$$
$$= \frac{1}{12}(3x+1)^4 + C.$$

Example 2.

$$\int \cos^3 x\,dx = \int \cos^2 x \cos x\,dx = \int (1-\sin^2 x)\,d\sin x$$
$$= -\sin x - \frac{\sin^3 x}{3} + C.$$

Example 3.

$$\int \frac{x\,dx}{(x^2+1)^2} = \frac{1}{2}\int (x^2+1)^{-2}\,d(x^2+1)$$
$$= \frac{1}{2}\,\frac{(x^2+1)^{-1}}{-1} + C = -\frac{1}{2(x^2+1)} + C.$$

Example 4.

$$\int \frac{dx}{x \ln x} = \int \frac{d \ln x}{\ln x} = \ln (\ln x) + C.$$

Example 5.

$$\int \cos^2 x\,dx = \int \frac{1+\cos 2x}{2}\,dx$$
$$= \frac{1}{2}\int dx + \frac{1}{4}\int \cos 2x\,d\,2x$$
$$= \frac{1}{2}\,x + \frac{1}{4}\sin 2x + C.$$

Example 6.

$$\int \tan^4 x\,dx = \int \tan^2 x\,\frac{\sin^2 x}{\cos^2 x}\,dx$$
$$= \int \tan^2 x\,\frac{1-\cos^2 x}{\cos^2 x}\,dx$$
$$= \int \tan^2 x\,d\tan x - \int \tan^2 x\,dx$$
$$= \frac{1}{3}\tan^3 x - \int \frac{1-\cos^2 x}{\cos^2 x}\,dx$$
$$= \frac{1}{3}\tan^3 x - \tan x + x + C.$$

It is sometimes useful to replace the initial argument by the function of a new argument.

Example 7. $\int \frac{dx}{x^2+a^2}$.

Solution. We set $x = ay$ and then

$$\int \frac{dx}{x^2+a^2} = \int \frac{a\,dy}{a^2y^2+a^2} = \frac{1}{a}\int \frac{dy}{y^2+1}$$
$$= \frac{1}{a}\arctan y + C = \frac{1}{a}\arctan \frac{x}{a} + C.$$

Example 8. $\int \sqrt{r^2-x^2}\,dx$.

Solution. We set $x = r \sin \varphi$ and then $dx = r \cos \varphi \, d\varphi$,

$$\begin{aligned}
&\int \sqrt{r^2 - x^2}\, dx \\
&= \int \sqrt{r^2 - r^2 \cdot \sin^2 \varphi}\, r \cos \varphi \, d\varphi = r^2 \int \cos^2 \varphi \, d\varphi \\
&= \frac{r^2}{2} \int \left(1 + \cos 2\varphi\right) d\varphi = \frac{r^2 \varphi}{2} + \frac{r^2}{4} \int \cos 2\varphi \, d\, 2\varphi \\
&= \frac{r^2 \varphi}{2} + \frac{r^2}{4} \sin 2\varphi + C = \frac{r^2 \varphi}{2} + \frac{r^2}{2} \sin \varphi \cos \varphi + C \\
&= \frac{r^2 \varphi}{2} + \frac{r^2}{2} \sin \varphi \sqrt{1 - \sin^2 \varphi} + C \\
&= \frac{r^2 \varphi}{2} + \frac{1}{2} r \sin \varphi \sqrt{r^2 - r^2 \cdot \sin^2 \varphi} + C \\
&= \frac{r^2}{2} \arcsin \frac{x}{r} + \frac{1}{2} x \sqrt{r^2 - x^2} + C.
\end{aligned}$$

◀ 6.5. Integration by Parts

The technique indicated in the title is based on the formula for the differentiation of the product $d(uv) = u\, dv + v\, du$, or $u\, dv = d(uv) - v\, du$. From the last relation it follows that

$$\int u\, dv = uv - \int v\, du.$$

Thus this technique makes it possible to reduce the calculation of one integral to that of another. The factors u, dv in the element of integration must be chosen such that the new integral would be simpler than the initial one.

Example 1.

$$\begin{aligned}
\int x \cos x \, dx &= \int x\, d \sin x = x \sin x - \int \sin x \, dx \\
&= x \sin x + \cos x + C.
\end{aligned}$$

Example 2.

$$\begin{aligned}
\int \ln x \, dx &= x \ln x - \int x\, d \ln x \\
&= x \ln x - \int x \cdot \frac{dx}{x} = x \ln x - x + C.
\end{aligned}$$

Example 3.

$$\int x^3 e^x\,dx = \int x^3\,de^x = x^3e^x - \int e^x 3x^2\,dx$$

$$= x^3e^x - 3\int x^2\,de^x = x^3e^x - 3x^2e^x + 3\int e^x\,2x\,dx$$

$$= x^3e^x - 3x^2e^x + 6\int x\,de^x$$

$$= x^3e^x - 3x^2e^x + 6xe^x - 6\int e^x\,dx$$

$$= x^3e^x - 3x^2e^x + 6xe^x - 6e^x + C$$

$$= (x^3 - 3x^2 + 6x - 6)\,e^x + C. \blacktriangleright$$

6.6. The Area of a Curvilinear Trapezoid

Consider a function $y = f(x)$ continuous on the interval $[a, b]$. First we assume that all the values of the function are nonnegative. Then the graph of the function $f(x)$ lies in the upper half-plane (Fig. 6.1). Let us calculate the area

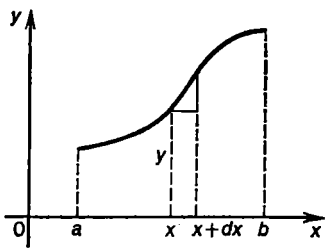


Fig. 6.1

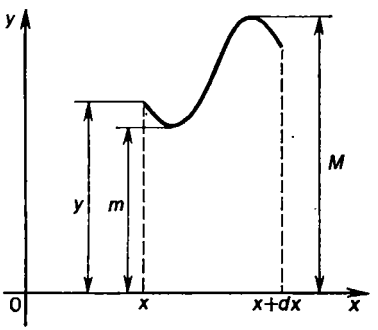


Fig. 6.2

of the part of the plane bounded by the x-axis, the straight lines $x = a$ and $x = b$ and the graph of the function $y = f(x)$. This figure is known as a *curvilinear trapezoid.* We consider the required area to be the value of the function $F(x)$, for $x = b$, which expresses the area of the curvilinear trapezoid with a variable right side which has an abscissa x. If we displace the right side a little so that x gets an incre-

ment dx, then the area gets an increment $\Delta F(x)$ equal to the area of a narrow strip. This area differs from the area $y\,dx$ of the rectangular strip by a small quantity which tends to zero appreciably faster than dx. Indeed, both $\Delta F(x)$ and $y\,dx$ are contained in a strip with altitude M and contain a strip with altitude m, where M and m are the greatest and the least value of $f(x)$ on the interval $[x, x+dx]$ (Fig. 6.2). Therefore $\Delta F(x) - y\,dx$ does not exceed $(M-m)\,dx$ in absolute value. But $M-m \to 0$ as $dx \to 0$ by virtue of the continuity of $f(x)$, i.e. $\Delta F(x) - y\,dx$ tends to zero appreciably faster than dx. Thus $y\,dx$ is the principal part of $\Delta F(x)$ and the product $y\,dx$ is linear with respect to dx. According to the fundamental property of the differential, this means that $y\,dx = dF(x)$. Thus $F(x)$ is an antiderivative of the differential $y\,dx$. Let $F_0(x)$ be some antiderivative. Then $F(x) = F_0(x) + C$. We can find the value of the constant C taking into account the fact that the area of $F(x)$ vanishes for $x = a$. Therefore $0 = F(a) = F_0(a) + C$, whence $C = -F_0(a)$, $F(x) = F_0(x) - F_0(a)$ and the required area

$$F(b) = F_0(b) - F_0(a).$$

The difference $F_0(b) - F_0(a)$, where F_0 is an antiderivative of the function $f(x)$ continuous on $[a, b]$, is the *definite integral* of $f(x)\,dx$ and is designated as

$$\int_a^b f(x)\,dx.$$

This term and this designation are also accepted by definition for $a > b$, the only requirement being that the function $f(x)$ should be continuous on the interval $[b, a]$.

The value of the definite integral does not depend on the choice of the antiderivative $F_0(x)$ since every antiderivative $F_1(x)$ differs from $F_0(x)$ only by a constant term and so

$$F_1(b) - F_1(a) = (F_0(b) + C) - (F_0(a) + C) = F_0(b) - F_0(a).$$

Thus, if $a < b$, the function $f(x)$ is continuous on $[a, b]$ and does not assume negative values, then the integral $\int_a^b f(x)\,dx$ is equal to the area of a curvilinear trapezoid

bounded by the x-axis, the straight lines $x = a$ and $x = b$ and the graph of the function $f(x)$.

We shall give now a geometric interpretation of the integral $\int_a^b f(x)\,dx$ for $a < b$ of the function $f(x)$, continuous on $[a, b]$, which can assume positive as well as negative values. The graph of the function $f(x)$ may pass from the

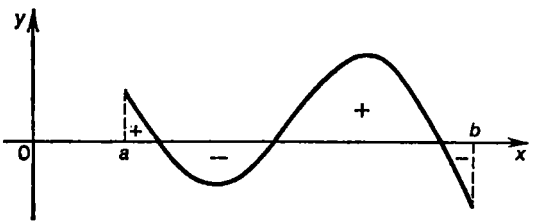


Fig. 6.3

upper half-plane to the lower one and back again. Let us consider a figure bounded by the graph of the function $f(x)$, the x-axis and the straight lines $x = a$ and $x = b$. The x-axis cuts this figure into parts some of which, corresponding to the positive values of $f(x)$, lie in the upper half-plane and the other parts, corresponding to the negative values of $f(x)$, lie in the lower half-plane (Fig. 6.3).

We introduce into consideration the algebraic sum of the areas of these parts taking the areas of the parts lying in the lower half-plane with the minus sign. As in the case of a curvilinear trapezoid, we first consider the quantity $F(x)$ for the figure with a variable right side and find its differential, i.e. the principal part of the increment, linear with respect to dx. If, for a certain x, we have $y = f(x) > 0$, then $dF(x) = y\,dx$, just as in the case of a curvilinear trapezoid. And if $y = f(x) < 0$, then for $dx > 0$ the quantity in question will assume a negative increment whose principal part is equal to $|y|\,dx$ in absolute value and algebraically it is again equal to $y\,dx$. Thus, in the case in question, the quantity $F(x)$ also proves to be an antiderivative of $f(x)$. From this we infer, as in the case of $f(x) \geqslant 0$, that $F(x) = F_0(x) - F_0(a)$, where $F_0(x)$ is any antiderivative of $f(x)$

and, finally,

$$F(b) = F_0(b) - F_0(a) = \int_a^b f(x)\,dx.$$

We have thus arrived at the following geometric interpretation of the definite integral of a continuous function. For $a < b$, the integral $\int_a^b f(x)\,dx$ is equal to the algebraic sum of the areas of the parts of the figure bounded by the graph of the function $y = f(x)$, the x-axis and the straight lines $x = a$ and $x = b$, the areas of the parts lying above the x-axis being taken with the plus sign and those of the areas lying below the x-axis, being taken with the minus sign.

We have also obtained an essentially important result concerning the existence of an antiderivative for any continuous function $y = f(x)$. A similar algebraic sum of the areas of the figure described above with a variable right side and with abscissa x, i.e. a definite integral with a variable upper limit x is exactly such an antiderivative. The notation $\int_a^x f(x)\,dx$ is not really good since the same letter x denotes the upper limit in the integral and the variable with respect to which the integration is carried out. But the value of the integral is evidently independent of the designation of the integration variable, i.e. for example,

$$\int_a^b f(x)\,dx = \int_a^b f(z)\,dz.$$

Therefore, when writing a definite integral with a variable upper limit x, it is expedient to change the designation of the integration variable, say, write it in the form

$$\int_a^x f(z)\,dz.$$

Thus, taking the existence of the areas we spoke of above on trust, we have proved the existence of an antiderivative of the function $F(x)$ for any continuous function $f(x)$ as an algebraic sum of the areas of the above figure with a variable

right-hand bound. Such an antiderivative can be written as

$$\int_a^x f(z)\,dz.$$

This result is not at all evident from the point of view of the technique of integration. The matter is that there are integrals that cannot be "taken", i.e. such that their elements of integration do not have antiderivatives which can be expressed in terms of elementary functions that we know.

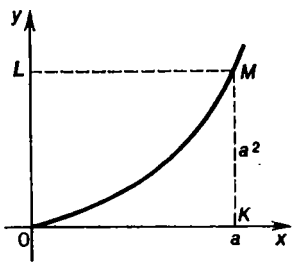


Fig. 6.4

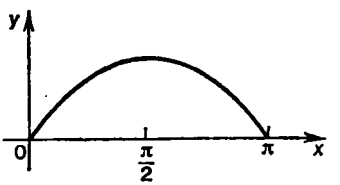


Fig. 6.5

Antiderivatives of this kind are new functions which include some functions important for applications. The "probability integral" $\int_0^x e^{-t^2}\,dt$, the "integral logarithm" $\int_2^x \frac{dz}{\ln z}$, the "integral sine" $\int \frac{\sin z}{z}\,dz$ and many others are examples of these new functions.

Let us consider some examples of calculating areas.

Example 1. Find the area S of a curvilinear triangle bounded by the abscissa axis, the parabola $y = x^2$ and the straight line $x = a$ (Fig. 6.4).

Solution.

$$S = \int_0^a x^2\,dx = \frac{x^3}{3}\Big|_{x=a} - \frac{x^3}{3}\Big|_{x=0} = \frac{a^3}{3} = \frac{1}{3}\,a \cdot a^2.$$

The required area is equal to one-third of the area of the rectangle $OKML$.

Example 2. Find the area bounded by the x-axis and one half-wave of the sine curve $y = \sin x$ (Fig. 6.5).

Solution.

$$S=\int_0^{\pi}\sin x\,dx=-\cos x|_{x=\pi}-(-\cos x)|_{x=0}=1-(-1)=2.$$

Example 3. Calculate $\int_0^{2\pi}\sin x\,dx$ and interpret the result you will obtain.

Solution.

$$\int_0^{2\pi}\sin x\,dx=-\cos x|_{x=2\pi}-(-\cos x)|_{x=0}$$

$$=-1-(-1)=0.$$

On the interval $[0, 2\pi]$ the sine curve makes two half-waves, the first of which lies above the x-axis and the other

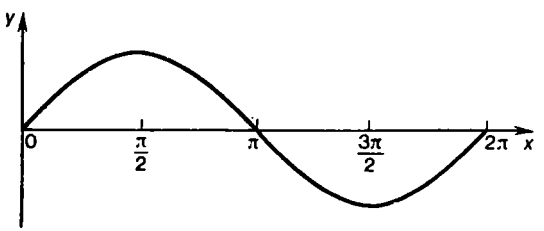


Fig. 6.6

lies below it (Fig. 6.6). The areas bounded by them are equal and the corresponding terms which appear in the geometric representation of the definite integral cancel each other.

Example 4. Find the area bounded by the x-axis, the hyperbola $y=1/x$ and the straight lines $x=b$ and $x=ab$, $a>1$, $b>0$.

Solution.

$$S=\int_b^{ab}\frac{1}{x}\,dx=\ln x|_{x=ab}-\ln x|_{x=b}=\ln ab-\ln b=\ln a.$$

For $b=1$ we get $S=\int_1^a\frac{1}{x}\,dx=\ln a$.

We have arrived at an interesting geometric representation of a natural logarithm in the form of the area of a curvilinear trapezoid bounded by the x-axis, the graph of $y = 1/x$ and the straight lines $x = 1$ and $x = a$.

From this we can easily get a graphic interpretation of the fact that the logarithm of the product is equal to the

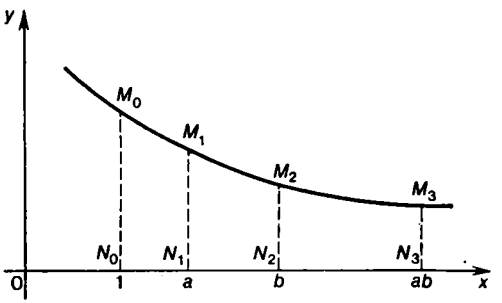


Fig. 6.7

sum of the logarithms (Fig. 6.7), namely, area $M_0N_0N_3M_3 =$ area $M_0N_0N_2M_2 +$ area $M_2N_2N_3M_3$. But area $M_0N_0N_3M_3 = \ln ab$, area $M_0N_0N_2M_2 = \ln b$, and we have calculated that area $M_2N_2N_3M_3 = \ln a$.

◀ We can verify in terms of geometry that area $M_2N_2N_3M_3 =$ area $M_0N_0N_1M_1$. Let us transform the plane shown in Fig. 6.7 by extending it b times along the x-axis and contrtating it also b times along the y-axis. Under this transformacion, the hyperbola $y = \frac{1}{x}$ will "slide" along itself since if $y = \frac{1}{x}$, then $\frac{y}{b} = \frac{1}{xb}$, i.e. $y_1 = \frac{1}{x_1}$, where $y_1 = \frac{y}{b}$, $x_1 = xb$. Therefore under this transformation, the figure $M_0N_0N_1M_1$ turns into a figure $M_2N_2N_3M_3$. But the areas of the figures do not change under this transformation, they increase b times upon the extension along the x-axis and decrease b times upon the contraction along the y-axis, i.e. return to the initial values.

If we did not know integrals and logarithms, we would set $\ln a =$ area $M_0N_0N_1M_1$ by definition and arrive at the fundamental property of logarithms:

$$\ln ab = \ln a + \ln b. \blacktriangleright$$

Example 5. Use integration to find the area of a circle of radius r.

Solution. A quarter of the area is equal to $\int_0^r \sqrt{r^2 - x^2}\, dx$ (Fig. 6.8). This integral is "rather difficult". We considered it in Example 8 in 6.4 and obtained

$$\int \sqrt{r^2 - x^2}\, dx = \frac{r^2}{2} \arcsin \frac{x}{r} + \frac{1}{2} x \sqrt{r^2 - x^2} + C,$$

whence

$$\int_0^r \sqrt{r^2 - x^2}\, dx = \left(\frac{r^2}{2} \arcsin \frac{x}{r} + \frac{1}{2} x \sqrt{r^2 - x^2} \right)_{x=r} - \left(\frac{r^2}{2} \arcsin \frac{x}{r} + \frac{1}{2} x \sqrt{r^2 - x^2} \right)_{x=0} = \frac{r^2}{2} \cdot \frac{\pi}{2} = \frac{\pi r^2}{4}$$

and so the area of the whole circle is πr^2.

We see that in this elementary case the problem of finding the area was solved by means of a "difficult integral".

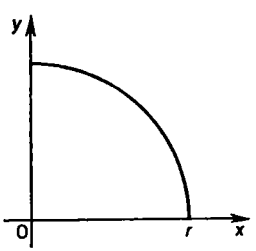


Fig. 6.8

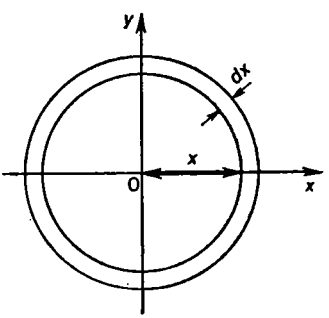


Fig. 6.9

Because of this fact, the method of calculating areas by means of integrals is less efficient than the elementary methods.

There are other techniques of representing the area of a circle as one of the values of the variable area. We shall show two of these techniques.

We consider a circle of a variable radius x (Fig. 6.9). Assume that its area is $S(x)$. An infinitely small increment $\Delta S(x)$ which appears when x increases by dx is equal to the area of a narrow strip between the circle of radius x and that of radius $x + dx$. The width of this strip is dx and we can take $2\pi x$ as its length when we calculate the principal part of the increment of the area. Thus $dS(x) = 2\pi x\,dx$, whence it follows that

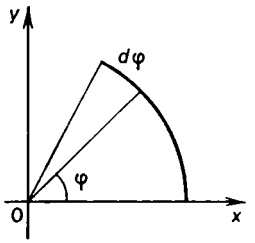


Fig. 6.10

$$S = \int_0^r 2\pi x\,dx = 2\pi \frac{x^2}{2}\Big|_{x=r} - 2\pi \frac{x^2}{2}\Big|_{x=0} = \pi r^2.$$

Here is one more technique. Consider a sector with a variable angle φ (Fig. 6.10). The increment of the area $S(\varphi)$ of the sector has the area of the triangle with base $r\,d\varphi$ and altitude r as its principal part. It is equal to $\frac{1}{2} r^2\,d\varphi$. Thus $dS(\varphi) = \frac{1}{2} r^2\,d\varphi$ and the area of the circle is

$$\frac{1}{2}\int_0^{2\pi} r^2\,d\varphi = \frac{1}{2} r^2 (\varphi|_{\varphi=2\pi} - \varphi|_{\varphi=0}) = \pi r^2.$$

6.7. Basic Properties of the Definite Integral

Consider the basic properties of the definite integral:

1. $\int_a^b (f_1(x) + f_2(x))\,dx = \int_a^b f_1(x)\,dx + \int_a^b f_2(x)\,dx.$

2. $\int_a^b mf(x)\,dx = m\int_a^b f(x)\,dx$

(m is a constant).

These properties follow immediately from similar properties of indefinite integrals.

3. *Let a, b, c belong to the interval on which the function f (x) is continuous. Then*

$$\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx.$$

Proof. Indeed, assume that $F_0(x)$ is an antiderivative of the function $f(x)\,dx$. Then

$$\int_a^c f(x)\,dx + \int_c^b f(x)\,dx = F_0(c) - F_0(a) + F_0(b) - F_0(c)$$

$$= F_0(b) - F_0(a) = \int_a^b f(x)\,dx.$$

4. It is evident that $\int_a^b f(x)\,dx = -\int_b^a f(x)\,dx$.

Properties 5, 6 and 7 that follow are geometrically almost obvious, but we shall prove them all the same, proceeding from the definition.

5. *Let $a < b$ and $f(x) > 0$ at all points of $[a, b]$. Then* $\int_a^b f(x)\,dx > 0$.

Proof. Indeed, assume that $F_0(x)$ is an antiderivative. It increases since its derivative $F_0'(x) = f(x)$ is positive. Therefore $F_0(a) < F_0(b)$ and $\int_a^b f(x)\,dx = F_0(b) - F_0(a) > 0$.

6. *Let $a < b$ and $f(x) \geqslant 0$ at all points of $[a, b]$. Then*

$$\int_a^b f(x)\,dx \geqslant 0.$$

Proof. Assume that ε is a positive number. Then $f(x) + \varepsilon > 0$ and, by virtue of Property 5,

$$\int_a^b (f(x) + \varepsilon)\,dx = \int_a^b f(x)\,dx + \varepsilon(b-a) > 0,$$

whence

$$\int_a^b f(x)\,dx > -\varepsilon(b-a).$$

But we can take ε arbitrarily small, and therefore

$$\int_a^b f(x)\,dx \geqslant 0.$$

7. *Let $a < b$ and $f(x)$ be continuous and nonnegative at all points of the interval $[a, b]$ and $f(x_0) > 0$ at least at one point x_0 within this interval. Then* $\int_a^b f(x)\,dx > 0$.

Proof. Since $f(x)$ is continuous, there is an interval $[x_0 - \delta, x_0 + \delta]$, belonging to $[a, b]$ in which $f(x)$ is positive at all points. We divide the integration interval into three parts:

$$\int_a^b f(x)\,dx = \int_a^{x_0-\delta} f(x)\,dx + \int_{x_0-\delta}^{x_0+\delta} f(x)\,dx + \int_{x_0+\delta}^b f(x)\,dx.$$

The first and the third term are nonnegative by virtue of Property 6. The second term is strictly positive by virtue of Property 5. Consequently,

$$\int_a^b f(x)\,dx > 0.$$

8. **The first mean value theorem for integrals.** *Let $a < b$ and $f(x)$ be continuous on $[a, b]$. Then there is a point c within $[a, b]$ for which*

$$\int_a^b f(x)\,dx = (b-a)\,f(c).$$

Proof. Let M and m be the greatest and the least value of $f(x)$ on $[a, b]$ and assume that $f(x_1) = M$, $f(x_2) = m$. Then $\int_a^b (M - f(x))\,dx \geqslant 0$, whence it follows that

$$\int_a^b f(x)\,dx \leqslant \int_a^b M\,dx = M(b-a).$$

Furthermore, $\int_a^b (f(x)-m)\,dx \geqslant 0$, whence

$$\int_a^b f(x)\,dx \geqslant \int_a^b m\,dx = m(b-a).$$

It follows from these inequalities that $m \leqslant \frac{1}{b-a}\int_a^b f(x)\,dx \leqslant M$. Thus the value $\frac{1}{b-a}\int_a^b f(x)\,dx$ is included between the least and the greatest value of $f(x)$. Since the function $f(x)$ is continuous, this value is equal to $f(c)$, where c is a point lying between x_1 and x_2 so that $x_1 < c < x_2$ if $x_1 < x_2$ or $x_2 < c < x_1$ if $x_2 < x_1$. The points x_1 and x_2 lie on the interval $[a, b]$ and consequently the point c is within this interval, i.e. $a < c < b$. Thus

$$\frac{1}{b-a}\int_a^b f(x)\,dx = f(c) \quad \text{for} \quad a < c < b$$

and

$$\int_a^b f(x)\,dx = (b-a)\,f(c),$$

and this is what we wished to prove.

Remark. The first mean value theorem for integrals is equivalent to the first mean value theorem for derivatives considered in 3.2 (Lemma 2) when applied to an antiderivative of $f(x)\,dx$. Indeed, if $F_0(x)$ is an antiderivative, then the first mean value theorem is equivalent to the relation $F_0(b) - F_0(a) = (b-a)\,F'(c)$, i.e. to the first mean value theorem for the function $F_0(x)$.

In the definition $\int_a^b f(x)\,dx = F(b) - F(a)$ of the definite integral it is essential that the function $f(x)$ should be continuous on the interval $[a, b]$. If we do not take this into account, we shall get absurd results. Thus, for instance, the function $1/x^2$ is positive for all values of x. However,

if we formally calculate $\int_{-1}^{+1} \frac{1}{x^2}\,dx$, we get a negative number

$$\int_{-1}^{+1} \frac{1}{x^2}\,dx = -\frac{1}{x}\Big|_{x=1} - \left(-\frac{1}{x}\right)\Big|_{x=-1} = -1-1 = -2,$$

which seems to contradict Property 5. The matter is that the function $1/x^2$ has a discontinuity for $x = 0$ and $\int_{-1}^{+1} \frac{dx}{x^2}$ is simply meaningless.

6.8. Representing an Integral as a Sum

Assume that $a < b$ and the function $f(x)$ is continuous on $[a, b]$. We divide the interval $[a, b]$ into n parts:

$$[x_0, x_1], \quad [x_1, x_2], \ \ldots, \ [x_{n-1}, x_n],$$

where

$$x_0 = a, \ x_n = b, \ x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n.$$

Then

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + \ldots + \int_{x_{n-1}}^{x_n} f(x)\,dx.$$

We apply the first mean value theorem to each of the terms. Then we obtain

$$\int_a^b f(x)\,dx = (x_1 - x_0) f(c_1) + (x_2 - x_1) f(c_2) + \ldots + (x_n - x_{n-1}) f(c_n).$$

Here $c_1, c_2, \ldots, c_n$ are definite points taken, respectively, from the intervals $[x_0, x_1]$, $[x_1, x_2]$, $\ldots$, $[x_{n-1}, x_n]$. The differences $x_1 - x_0, x_2 - x_1, \ldots, x_n - x_{n-1}$ can be regarded as the increments of the variable x when we pass from x_0 to x_1, from x_1 to x_2, $\ldots$, from x_{n-1} to x_n. Designating these

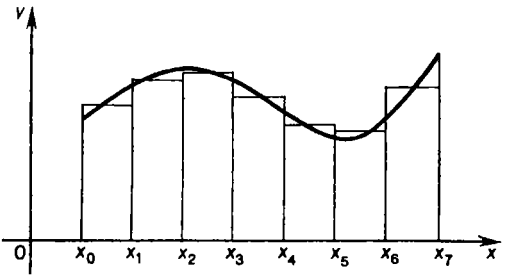


Fig. 6.11

increments as d_1x, d_2x, . . ., d_nx, we arrive at a relation

$$\int_a^b f(x)\,x = f(c_1)\,d_1x + f(c_2)\,d_2x + \ldots + f(c_n)\,d_nx.$$

This formula has a clear geometric meaning. To make it more graphic, we set $f(x) > 0$ (Fig. 6.11). In the figure $n = 7$. Every term $f(c_k)\,d_kx$ is equal to the area of a rectangle with base d_kx and with the altitude such that the area of the rectangle is equal to the area of the corresponding curvilinear trapezoid.

The formula we have derived explains, to some extent, the origin of the designation of the integral. We have mentioned above that the sign $\int$ is an extended letter S which is the first letter of the word "sum". The terms being summed up have the form of the products of the values $f(c_i)$ of the integrand function by the differentials of the independent variable. The limits a and b in the integral show what interval is divided into parts and, thus, from what interval all the values $f(c_i)$ of the function $f(x)$ are taken.

6.9. The Integral as the Limit of a Sum

The formula

$$\int_a^b f(x)\,dx = f(c_1)\,d_1x + f(c_2)\,d_2x + \ldots + f(c_n)\,d_nx$$

considered in the preceding section is exact but contains an essential indeterminateness: we do not know the exact values

$c_1, c_2, \ldots, c_n$ but only know to what intervals they belong. Now if instead of the values of $f(x)$ at the points $c_1, c_2, \ldots, c_n$ we take the values at some other points $z_1, z_2, \ldots, z_n$ from the same intervals, we get a sum

$$f(z_1)\, d_1x + f(z_2)\, d_2x + \ldots + f(z_n)\, d_nx$$

which is no longer exactly equal to $\int_a^b f(x)\, dx$ but, what is geometrically obvious (Fig. 6.12), differs but little from the integral $\int_a^b f(x)\, dx$, and this difference is the smaller, the

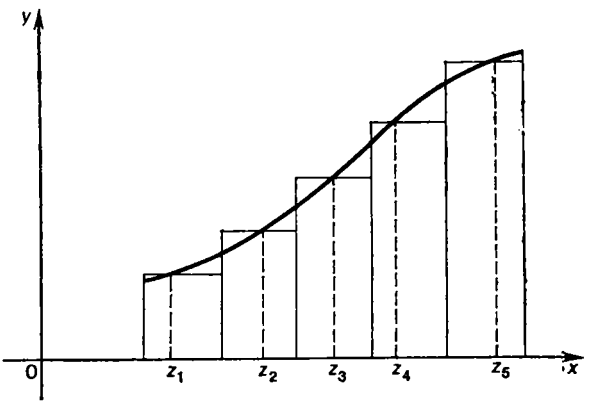


Fig. 6.12

finer the parts into which the integration interval is divided. The sum constructed in this way is known as an *integral sum* or the *Riemann sum.*

Among integral sums we distinguish an *upper integral sum*, in which we take the points of maximum of $f(x)$ on $[x_{i-1}, x_i]$ as z_i, and a *lower integral sum*, in which we take the points of minimum of $f(x)$ on $[x_{i-1}, x_i]$ as z_i. Thus the upper integral sum is

$$S_{\max} = M_1\, d_1x + M_2\, d_2x + \ldots + M_n\, d_nx,$$

where $M_1, M_2, \ldots, M_n$ are the greatest values of $f(x)$ on the intervals $[x_0, x_1]$, $[x_1, x_2]$, $\ldots$, $[x_{n-1}, x_n]$, and the lower

integral sum

$$S_{\min} = m_1 d_1x + m_2 d_2x + \ldots + m_n d_nx,$$

where $m_1, m_2, \ldots, m_n$ are the least values of $f(x)$ on the corresponding intervals. The geometric representation of the upper and lower integral sums is given in Figs. 6.13 and 6.14.

It is clear that any upper integral sum is greater than the integral (if $f(x)$ is nonconstant) and any lower sum is

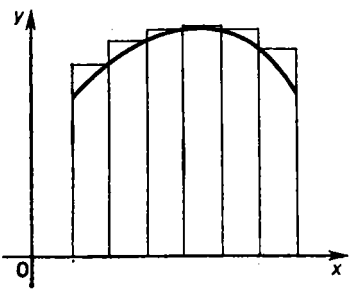


Fig. 6.13

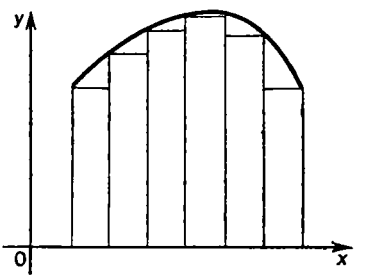


Fig. 6.14

smaller than the integral. And, for the given division, any integral sum is not greater than the upper sum and not smaller than the lower sum.

Let us estimate the difference between the upper and the lower integral sum for the same division of an integration interval. For this purpose, we designate the greatest of the differences $M_1 - m_1, M_2 - m_2, \ldots, M_n - m_n$ as δ. It is possible although not so easy to prove that δ can become arbitrarily small if we make the division intervals sufficiently small. We have

$$\begin{aligned} S_{\max} - S_{\min} &= (M_1 - m_1) d_1x + (M_2 - m_2) d_2x \\ &\quad + \ldots + (M_n - m_n) d_nx \\ &\leqslant \delta d_1x + \delta d_2x + \ldots + \delta d_nx \\ &= \delta (d_1x + d_2x + \ldots + d_nx) \\ &= \delta (x_1 - x_0 + x_2 - x_1 + \ldots + x_n - x_{n-1}) \\ &= \delta (x_n - x_0) = \delta (b - a). \end{aligned}$$

Thus S_{max} and S_{min} approach each other indefinitely when we make the subdivisions of the interval $[a, b]$ smaller and smaller, and, approaching each other, they approach the integral $\int_a^b f(x)\,dx$ which lies between S_{max} and S_{min} for any division. Thus S_{max} and S_{min} have a limit $\int_a^b f(x)\,dx$ on the assumption that we consider a sequence of smaller and smaller subintervals. Since any integral sum lies between S_{max} and S_{min}, any integral sum also has a limit $\int_a^b f(x)\,dx$ for an infinite refinement of the division of the interval $[a, b]$.

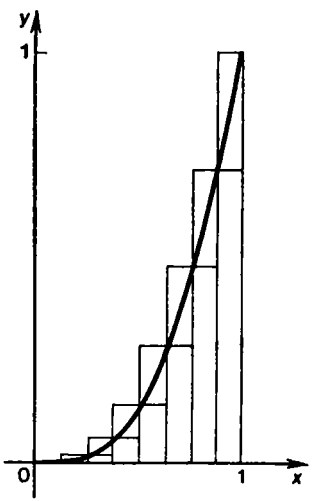


Fig. 6.15

Example 1. Calculate the upper and the lower integral sum for $\int_0^1 x^3\,dx$ proceeding from the division of the interval $[0, 1]$ into n equal parts.

Solution. The function $y = x^3$ increases and therefore on each interval $[x_{i-1}, x_i]$ the greatest value is attained at the right endpoint of the interval and the least value is attained at the left endpoint (Fig. 6.15). Here $x_0 = 0$, $x_1 = \frac{1}{n}$, $x_2 = \frac{2}{n}$, ..., $x_{n-1} = \frac{n-1}{n}$, $x_n = \frac{n}{n} = 1$, $d_1x = d_2x = \ldots = d_nx = \frac{1}{n}$. The lower sum

$$S_{min} = 0 + \left(\frac{1}{n}\right)^3 \frac{1}{n} + \left(\frac{2}{n}\right)^3 \frac{1}{n} + \ldots + \left(\frac{n-1}{n}\right)^3 \frac{1}{n}$$

$$= \frac{1^3 + 2^3 + \ldots + (n-1)^3}{4}.$$

The upper sum

$$S_{\max} = \left(\frac{1}{n}\right)^3 \frac{1}{n} + \left(\frac{2}{n}\right)^3 \frac{1}{n} + \ldots + \left(\frac{n}{n}\right)^3 \frac{1}{n}$$
$$= \frac{1^3 + 2^3 + \ldots + n^3}{n^4}.$$

It is clear that $S_{\max} - S_{\min} = \frac{n^3}{n^4} = \frac{1}{n} \to 0$ as $n \to \infty$.

To calculate the sums, recall that $1^3 + 2^3 + \ldots + n^3 = \frac{n^2 (n+1)^2}{4}$ and so

$$1^3 + 2^3 + \ldots + (n-1)^3 = \frac{n^2 (n-1)^2}{4}.$$

Consequently,

$$S_{\max} = \frac{1}{n^4} \cdot \frac{n^2 (n+1)^2}{4} = \frac{n^2 + 2n + 1}{4n^2}$$
$$= \frac{1}{4}\left(1 + \frac{2}{n} + \frac{1}{n^2}\right) \to \frac{1}{4} \text{ as } n \to \infty.$$

Correspondingly,

$$S_{\min} = \frac{1}{n^4} \cdot \frac{n^2 (n-1)^2}{4} = \frac{n^2 - 2n + 1}{4n^2}$$
$$= \frac{1}{4}\left(1 - \frac{2}{n} + \frac{1}{n^2}\right) \to \frac{1}{4} \text{ as } n \to \infty.$$

These sums must have a limit

$$\int_0^1 x^3 \, dx = \frac{x^4}{4}\Big|_{x=1} - \frac{x^4}{4}\Big|_{x=0} = \frac{1}{4},$$

and they have it indeed.

◀ In conclusion consider one more geometric interpretation of an integral sum

$$S = f(z_1)(x_1 - x_0) + f(z_2)(x_2 - x_1)$$
$$+ \ldots + f(z_n)(x_n - x_{n-1}),$$

where $a = x_0 < x_1 < x_2 < \ldots < x_{n-1} < x_n = b$ and $x_{i-1} \leqslant z_i \leqslant x_i$.

Assume that $F(x) = F_0(x) - F_0(a)$ is an antiderivative of $f(x)$ which becomes zero for $x = a$. Let us consider the following approximate representation of the graph of $F(x)$ made with the use of a broken line. We construct a family

of vertical straight lines $x = x_0$, $x = x_1$, $x = x_2$, . . ., $x = x_{n-1}$, $x = x_n$ (Fig. 6.16) and also a broken line with vertices at these straight lines. We shall do it in the following way. Proceeding from the point A_0 $(a, 0)$ through which the graph of $F(x)$ passes, we construct a segment of the straight line A_0A_1, with a slope $f(z_1) = F'(z_1)$, to the point where it

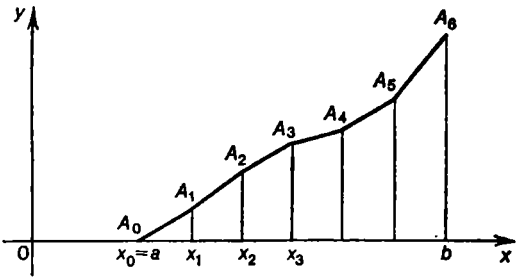


Fig. 6.16

meets the straight line $x = x_1$. This is the slope of the tangent to the graph of $F(z)$ at the point z which lies between x_0 and x. Therefore the segment A_0A_1 adjoins the corresponding part of the graph very closely. The ordinate of the point a_1 is evidently equal to $f(z_1)(x_1 - x_0)$. Then, from the point A_1 we draw a segment A_1A_2, with the slope $f(z_2) = F'(z_2)$, to the point where it meets the straight line $x = x_2$. This segment is parallel to the tangent to the graph of $F(x)$ at the point $x = z_2$. The ordinate of the point A_2 results from the ordinate of the point x_1 when the number $f(z_2)(x_2 - x_1)$ is added to it and so it is equal to $f(z_1)(x_1 - x_0) + f(z_2) \times (x_2 - x_1)$. Then we draw a segment A_2A_3, with a slope $f(z_3) = F'(z_3)$, to the point where it meets the straight line $x = x_3$ and so on. The ordinate of the point A_k results from the ordinate of the point A_{k-1} when the number $f(z_k)(x_k - x_{k-1})$ is added to it, and so it is equal to

$$f(z_1)(x_1 - x_0) + f(z_2)(x_2 - x_1) + \ldots + f(z_k)(x_k - x_{k-1}).$$

The ordinate of the last point A_n is equal to

$$f(z_1)(x_1 - x_0) + f(z_2)(x_2 - x_1) + \ldots + f(z_n)(x_n - x_{n-1}),$$

i.e. to the integral sum S. If we make the divisions finer and finer, then $\lim S = \int_a^b f(x)\,dx = F(b)$, i.e. the extreme right point of the constructed broken line with abscissa b converges to the number $F(b)$, i.e. to the point of the graph of $F(x)$ with abscissa b. It is easy to see that not only the extreme right point of the broken line approaches the extreme right point of the graph of $F(x)$, but also all the points of the broken line with a fixed abscissa will tend to the points of the graph of $F(x)$ with the same abscissas so that the constructed broken line indeed approaches the graph of $F(x)$ which becomes more and more accurate with an infinite division of the interval $[a, b]$.

The broken line corresponding to the upper integral sum lies entirely above the graph of $F(x)$ and the broken line corresponding to the lower sum lies entirely below this graph.

If we take c_i from the first mean value theorem as z_i,

$$\int_{x_{i-1}}^{x_i} f(x)\,dx = f(c_i)(x_i - x_{i-1}),$$

then the broken line will be inscribed in the graph of $F(x)$, i.e. its vertices will lie on this graph. ▶

6.10. Approximate Calculation of Integrals

The representation of an integral as the limit of an integral sum gives us a means for approximation of the integral. We know that by taking sufficiently small divisions of the integration interval we can get an integral sum arbitrarily close to the integral. If we simultaneously calculate the lower and the upper integral sum, we get approximations of the integral by defect and by excess. But we can indicate much more "economical" techniques of calculation than that of calculating lower and upper integral sums. Here are some of them.

1. The rule of rectangles. As an approximation of the integral we take an integral sum in which the values of the integrand are taken at the midpoints of the intervals into which the integration interval is divided. All these subintervals are assumed to have the same length equal to

$\frac{b-a}{n}$, where a and b are the limits of integration and n is the number of subintervals. Thus the integral sum has the form

$$S_{\text{rec}} = \frac{b-a}{n}\left[f\left(\frac{x_0+x_1}{2}\right)+f\left(\frac{x_1+x_2}{2}\right)+\ldots+f\left(\frac{x_{n-1}+x_n}{2}\right)\right]$$
$$= \frac{b-a}{n}(\bar{y}_1+\bar{y}_2+\ldots+\bar{y}_n),$$

where $\bar{y}_i$ denotes $f\left(\frac{x_{i-1}+x_i}{2}\right)$ (Fig. 6.17). Since $\frac{x_{i-1}+x_i}{2}$ lies midway between the straight lines $x = x_{i-1}$ and $x = x_i$

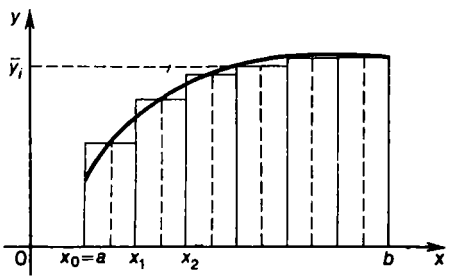


Fig. 6.17

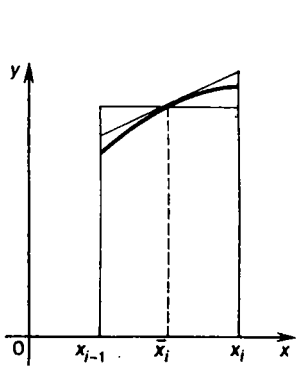


Fig. 6.18

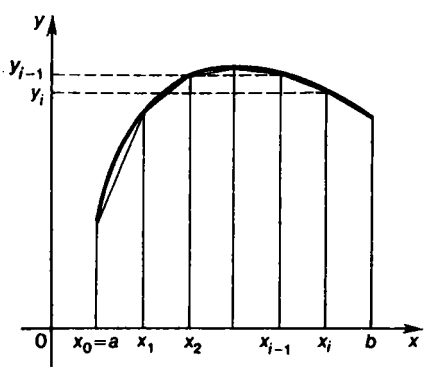


Fig. 6.19

which pass through the endpoints of the interval $[x_{i-1}, x_i]$, the area of the rectangle is equal to the area of the trapezoid (Fig. 6.18) bounded by the interval $[x_{i-1}, x_i]$, the straight

lines $x = x_{i-1}$ and $x = x_i$ and the tangent to the graph of $f(x)$ at the point with abscissa $\frac{x_{i-1}+x_i}{2} = \bar{x}_i$.

2. The trapezoidal rule. We divide the integration interval $[a, b]$ into n equal parts by the points $a = x_0, x_1, x_2, \ldots, x_{n-1}, x_n = b$. Then we replace the area of each curvilinear trapezoid between the straight lines $x = x_{i-1}$ and $x = x_i$ by the area of a rectilinear trapezoid bounded from above by the chord which connects the points of the graph with abscissas x_{i-1} and x_i. The bases of these trapezoids are $y_{i-1} = f(x_{i-1})$ and $y_i = f(x_i)$ (Fig. 6.19). The sum of the areas of these trapezoids is evidently

$$S_{\text{tr}} = \frac{b-a}{n}\left[\frac{y_0+y_1}{2} + \frac{y_1+y_2}{2} + \ldots + \frac{y_{n-1}+y_n}{2}\right]$$

$$= \frac{b-a}{n}\left(\frac{y_0}{2} + y_1 + \ldots + y_{n-1} + \frac{y_n}{2}\right).$$

3. The parabolic rule (Simpson's rule). This rule is based on the formula

$$S_{\text{par}} = \frac{1}{3}(S_{\text{tr}} + 2S_{\text{rec}})$$

◀ Let us substantiate it. Assume that we have a parabola $y = \varphi(x) = ax^2 + bx + c$. We shall seek the area of the curvilinear trapezoid bounded by the parabola $y = ax^2 + bx + c$, the straight lines $x = -h/2$ and $x = h/2$ (separated by the distance h) and a segment of the x-axis. This area is equal to

$$\int_{-h/2}^{h/2} (ax^2 + bx + c)\,dx = \frac{ah^3}{12} + ch = \frac{h}{3}\left(\frac{ah^2}{4} + 3c\right).$$

We express it in terms of $y_0 = \varphi\left(-\frac{h}{2}\right) = \frac{ah^2}{4} - \frac{bh}{2} + c$, $y_1 = \varphi\left(\frac{h}{2}\right) = \frac{ah^2}{4} + \frac{bh}{2} + c$ and $\bar{y}_1 = \varphi(0) = c$. We have $y_0 + y_1 = \frac{ah^2}{2} + 2c$, $\frac{y_0+y_1}{2} = \frac{ah^2}{4} + c$, and so

$$\int_{-h/2}^{h/2} (ax^2 + bx + c)\,dx = \frac{h}{3}\left(\frac{y_0+y_1}{2} + 2\bar{y}_1\right).$$

This formula remains valid for any parabolic trapezoid with altitude h since when it is displaced along the x-axis, the form of the equation for the parabola $y = ax^2 + bx + c$ remains the same and only the coefficients change.

We shall show now that through any three points (x_0, y_0), $(\bar{x}_1, \bar{y}_1)$ and (x_1, y_1) with pairwise different abscissas we can draw a parabola with an equation of the form $y = ax^2 + bx + c$. The requirement that the parabola should pass through the given points can be written as relations

$$y_0 = ax_0^2 + bx_0 + c, \quad \bar{y}_1 = a\bar{x}_1^2 + b\bar{x}_1 + c, \quad y_1 = ax_1^2 + bx_1 + c.$$

These relations constitute a system of linear equations with respect to the unknown coefficients a, b and c which is easy to solve. Subtracting the first equation from the second and the third, we obtain

$$\bar{y}_1 - y_0 = a\bar{x}_1^2 + b\bar{x}_1 - ax_0^2 - bx_0 = (\bar{x}_1 - x_0)(a(\bar{x}_1 + x_0) + b),$$
$$y_1 - y_0 = ax_1^2 + bx_1 - ax_0^2 - bx_0 = (x_1 - x_0)(a(x_1 + x_0) + b).$$

Dividing these relations by $\bar{x}_1 - x_0$ and, respectively, by $x_1 - x_0$, we obtain

$$\frac{\bar{y}_1 - y_0}{\bar{x}_1 - x_0} = a(\bar{x}_1 + x_0) + b, \quad \frac{y_1 - y_0}{x_1 - x_0} = a(x_1 + x_0) + b.$$

Subtracting the second relation from the first, we obtain

$$\frac{\bar{y}_1 - y_0}{\bar{x}_1 - x_0} - \frac{y_1 - y_0}{x_1 - x_0} = a(\bar{x}_1 - x_1).$$

We have thus reduced the system to the form

$$y_0 = ax_0^2 + bx_0 + c, \quad \frac{\bar{y}_1 - y_0}{\bar{x}_1 - x_0} = a(\bar{x}_1 + x_0) + b,$$
$$\frac{\bar{y}_1 - y_0}{\bar{x}_1 - x_0} - \frac{y_1 - y_0}{x_1 - x_0} = a(\bar{x}_1 - x_1).$$

From the last equation we find a, from the second we find b and from the first, c. All the necessary divisions are quite possible since x_0, $\bar{x}_1$ and x are pairwise distinct by the hypothesis. Thus, we can, indeed, draw a parabola $y = ax^2 + bx + c$ through the points (x_0, y_0), $(\bar{x}_1, \bar{y}_1)$ and (x_1, y_1) with pairwise distinct abscissas.

Let us return to the problem of approximate calculation of $\int_a^b f(x)\,dx$. We divide the integration interval into n equal parts and on each part $[x_{i-1}, x_i]$ replace the graph of the integrand function by a segment of the parabola which passes through the points of the graph with abscissas x_{i-1}, $\bar{x}_i = \frac{x_{i-1}+x_i}{2}$ and x_i. Clearly, such a segment of the parabola will, in general, differ but little from the corresponding part of the graph of the function. Replacing $\int_a^b f(x)\,dx$ by the sum of the areas of parabolic trapezoids, we obtain

$$\int_a^b f(x)\,dx \cong \frac{b-a}{n}\cdot\frac{1}{3}\left[\frac{y_0+y_1}{2}+2\bar{y}_1+\frac{y_1+y_2}{2}\right.$$

$$\left.+2\bar{y}_2+\ldots+\frac{y_{n-1}+y_n}{2}+2\bar{y}_n\right]$$

$$=\frac{1}{3}\cdot\frac{b-a}{n}\left[\frac{y_0}{2}+y_1+\ldots+y_{n-1}+\frac{y_n}{2}\right.$$

$$\left.+2(\bar{y}_1+\bar{y}_2+\ldots+\bar{y}_n)\right]=\frac{1}{3}(S_{\text{tr}}+2S_{\text{rec}}).$$

There are also other methods of approximate calculation of integrals. ▶

Example 1. Make approximate calculation of $\frac{\pi}{4}=\int_0^1 \frac{dx}{1+x^2}$.

Solution. For $n = 1$ we have

$$y_0=1,\quad \bar{y}_1=\frac{4}{5}=0.8,\quad y_1=\frac{1}{2},\quad S_{\text{rec}}=0.8,$$

$$S_{\text{tr}}=\frac{3}{4}=0.75,\quad S_{\text{par}}=\frac{8}{5}+\frac{3}{4}=\frac{47}{60}\cong 0.7833.$$

Actually, with four decimal digits $\frac{\pi}{4}$ is approximately equal to 0.7854.

For $n = 3$ we have

$$y_0 = 1, \quad y_1 = \frac{9}{10} = 0.9000000,$$

$$y_2 = \frac{9}{13} = 0.69230769, \quad y_3 = \frac{1}{2},$$

$$S_{\text{tr}} = 0.78076923,$$

$$y_1 = \frac{36}{37} = 0.97297297, \quad \bar{y}_1 = \frac{4}{5} = 0.8000000,$$

$$\bar{y}_2 = \frac{36}{61} = 0.59016393, \quad S_{\text{rec}} = 0.78771230,$$

$$S_{\text{par}} = \frac{S_{\text{tr}} + 2S_{\text{rec}}}{3} = 0.78539794.$$

We have made calculations with eight decimal digits. Actually, with eight decimal digits the value of $\frac{\pi}{4}$ is 0.78539816. For $n = 5$, when we make calculations with eight decimal digits, all the eight digits prove to be valid.

6.11. The Volume of a Solid of Revolution

Let $y = f(x)$ be a continuous nonnegative function defined on the interval $[a, b]$. Imagine a solid resulting from the rotation, about the x-axis, of a curvilinear trapezoid

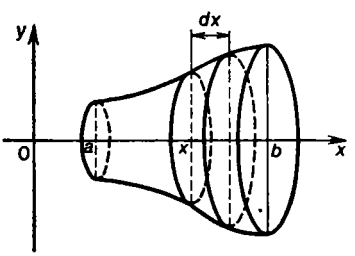


Fig. 6.20

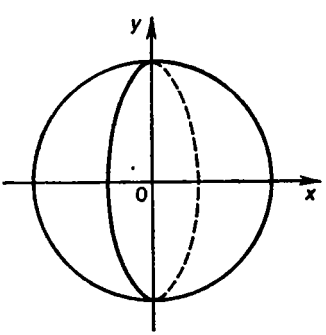


Fig. 6.21

bounded by the graph of the function $y = f(x)$, the x-axis and the straight lines $x = a$ and $x = b$ (Fig. 6 20). To calculate the volume of this solid, we shall consider a similar

solid with a variable right side which cuts the x-axis at the point x. The volume of this solid $V(x)$ is a function of x. For an infinitesimal (i.e. infinitely decreasing) increment dx the volume increases by the volume of an infinitely thin (i.e. becoming infinitely thin) layer of width dx with the base area $\pi f^2(x)$. Thus

$$dV(x) = \pi f^2(x)\,dx \quad \text{and} \quad V(x) = \int \pi f^2(x)\,dx = F_0(x) + C,$$

where $F_0(x)$ is an antiderivative of $\pi f^2(x)\,dx$. For $x = a$ we find that $0 = F(a) = F_0(a) + C$, whence $C = -F_0(a)$, the volume $V(x)$ is equal to $F_0(x) - F_0(a)$ and the required volume

$$V = V(b) = F_0(b) - F_0(a) = \int_a^b \pi y^2\,dx.$$

Example 1. Calculate the volume of a sphere of radius r.

Solution. The sphere results from the rotation of a semicircle of radius r about the diameter (Fig. 6.21). Assuming the diameter to be the x-axis, we find that the semicircle is bounded by a semicircumference with an equation $y = \sqrt{r^2 - x^2}$. We must perform integration from the abscissa of the left end of the semicircle to the abscissa of the right end, i.e.

$$V = \int_{-r}^{r} \pi(\sqrt{r^2 - x^2})^2\,dx = \int_{-r}^{r} \pi(r^2 - x^2)\,dx.$$

Making the necessary calculations, we obtain

$$V = \pi\left(r^2x - \frac{x^3}{3}\right)_{x=r} - \pi\left(r^2x - \frac{x^3}{3}\right)_{x=-r}$$

$$= \pi r^3 - \frac{\pi r^3}{3} + \pi r^3 - \frac{\pi r^3}{3} = \frac{4}{3}\pi r^3.$$

Example 2. Calculate the volume of a paraboloid of revolution, i.e. a solid resulting from the rotation, about the x-axis, of a "parabolic triangle" bounded by the upper half of the parabola $y^2 = cx$, the x-axis and the straight line $x = a$ (Fig. 6.22).

Solution.

$$V = \int_0^a \pi y^2\,dx = \int_0^a \pi cx\,dx = \pi c\,\frac{x^2}{2}\Big|_{x=a} - \pi c\,\frac{x^2}{2}\Big|_{x=0}$$

$$= \frac{\pi ca^2}{2} = \pi ca \cdot \frac{a}{2} = \pi\left(\sqrt{ca}\right)^2 \cdot \frac{a}{2} = \frac{Sa}{2}.$$

Here $S = \pi\left(\sqrt{ca}\right)^2$ is the area of a circle which bounds the paraboloid from the right. Thus the volume of the paraboloid

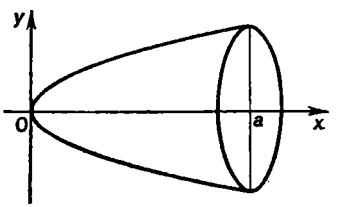


Fig. 6.22

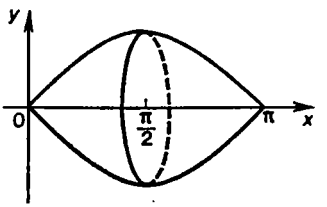


Fig. 6.23

of revolution is equal to half the volume of the cylinder circumscribed about it with the base area S and altitude a.

Example 3. Find the volume of a solid resulting from the rotation, about the x-axis, of a figure bounded by the x-axis and one half-wave of the sine curve $y = \sin x$ (Fig. 6.23).

Solution.

$$V = \int_0^\pi \pi \sin^2 x\,dx = \int_0^\pi \pi\,\frac{1-\cos 2x}{2}\,dx$$

$$= \pi\left(\frac{x}{2} - \frac{\sin 2x}{4}\right)_{x=\pi} - \pi\left(\frac{x}{2} - \frac{\sin 2x}{4}\right)_{x=0} = \frac{\pi^2}{2}.$$

6.12. The Length of the Arc of a Curve

Let $y = f(x)$ be a continuous and differentiable function on the interval $[a, b]$. We shall consider the problem of calculating the length l of the graph of $f(x)$ from the point with abscissa a to the point with abscissa b. We designate as $l(x)$ the length of the curve from the point with abscissa a to the point with abscissa x (Fig. 6.24). Assume that the abscissa x has got an infinitesimal increment dx. Then y

gets an infinitesimal increment Δy which differs from the increment dy along the tangent by an infinitesimal tending to zero appreciably faster than dx. The increment Δl of

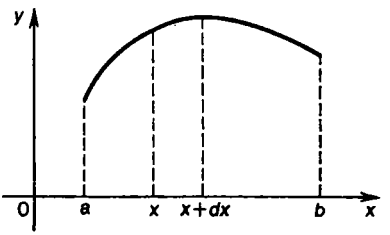


Fig. 6.24

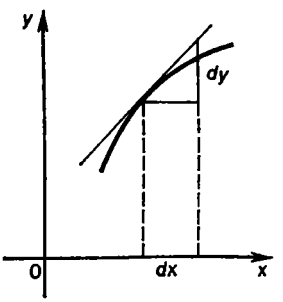


Fig. 6.25

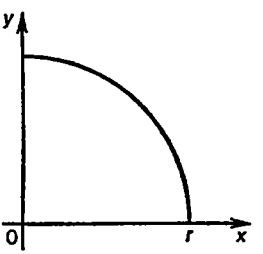


Fig. 6.26

the length of the curve differs from the length of the corresponding segment of the tangent

$$\sqrt{dx^2+dy^2}=\sqrt{1+(y')^2}\,dx$$

(Fig. 6.25) by an infinitesimal which is appreciably smaller than dx. The latter is linear with respect to dx and is the principal part of Δl. Consequently,

$$dl=\sqrt{1+(y')^2}\,dx \quad \text{and} \quad l=\int_a^b \sqrt{1+(y')^2}\,dx.$$

Example 1. By means of integration find the length of a quarter of a circle with radius r (Fig. 6.26).

Solution. Here $y=\sqrt{r^2-x^2}$, $a=0$, $b=r$,

$$y'=-\frac{2x}{2\sqrt{r^2-x^2}}=-\frac{x}{\sqrt{r^2-x^2}},$$

$$1+(y')^2=1+\frac{x^2}{r^2-x^2}=\frac{r^2-x^2+x^2}{r^2-x^2}=\frac{r^2}{r^2-x^2},$$

$$l=\int_0^r \frac{r}{\sqrt{r^2-x^2}}\,dx=\int_0^r \frac{1}{\sqrt{1-\frac{x^2}{r^2}}}\,dx$$

$$=r\int_0^r \frac{1}{\sqrt{1-\frac{x^2}{r^2}}}\,d\left(\frac{x}{r}\right)$$

$$=r\arcsin\frac{x}{r}\Big|_{x=r}-r\arcsin\frac{x}{r}\Big|_{x=0}=\frac{r\pi}{2}$$

and this is what we could expect.

Note that the calculation of the length of an ellipse reduces to the calculation of an integral that "cannot be taken", i.e. an integral which cannot be expressed in terms of elementary functions. The same refers to the calculation of the length of the arc of the hyperbola $y = 1/x$ and the length of the arc of a sine curve. The length of the arc of a parabola can be reduced to an integral which is rather complicated although it can be expressed in terms of elementary functions.

Example 2. Find the length of the arc of the half-cubic parabola $y = x^{3/2}$ from the point with abscissa 0 to the point with abscissa a.

Solution.

$$l=\int_0^a \sqrt{1+\frac{9}{4}x}\,dx=\frac{4}{9}\int_0^a \left(1+\frac{9}{4}x\right)^{1/2} d\left(1+\frac{9}{4}x\right)$$

$$=\frac{4}{9}\cdot\frac{\left(1+\frac{9}{4}x\right)^{3/2}}{3/2}\Big|_{x=a}-\frac{4}{9}\cdot\frac{\left(1+\frac{9}{4}x\right)^{3/2}}{3/2}\Big|_{x=0}$$

$$=\frac{8}{27}\left(1+\frac{9}{4}a\right)^{3/2}-\frac{8}{27}.$$

6.13. The Lateral Area of a Solid of Revolution

Assume that a solid has been obtained as a result of the rotation of a curvilinear trapezoid bounded by the x-axis, the straight lines $x = a$ and $x = b$ and the graph of the differentiable function $y = f(x) \geqslant 0$. We have to find the lateral area of this solid. We introduce into consideration

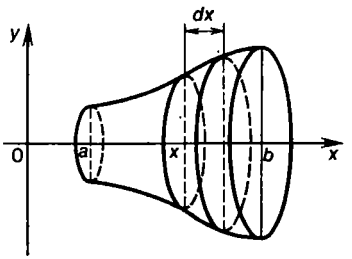


Fig. 6.27

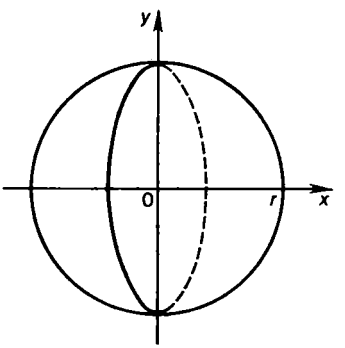


Fig. 6.28

the area $P(x)$ of a similar solid but bounded by a variable right side which cuts the x-axis at a point with abscissa x (Fig. 6.27).

For an infinitesimal increment dx, the principal part of the increment $\Delta P(x)$ is the area of a band of length $2\pi y$ and width dl so that $dP(x) = 2\pi y\, dl$ and

$$P = \int_a^b 2\pi y\, dl = \int_a^b 2\pi y \sqrt{1 + (y')^2}\, dx.$$

Example 1. Find the area of the surface of a sphere of radius r.

Solution. Here $y = \sqrt{r^2 - x^2}$, $a = -r$, $b = r$ (Fig. 6.28).

$$P = \int_{-r}^{r} 2\pi \sqrt{r^2 - x^2} \sqrt{1 + \left(\frac{-x}{\sqrt{r^2 - x^2}}\right)^2}\, dx$$

$$= \int_{-r}^{r} 2\pi \sqrt{r^2 - x^2} \cdot \frac{r}{\sqrt{r^2 - x^2}}\, dx = \int_{-r}^{r} 2\pi r\, dx$$

$$= 2\pi r x|_{x=r} - 2\pi r x|_{x=-r} = 4\pi r^2.$$

Example 2. Find the lateral area of the paraboloid discussed in Example in 6.11.

Solution. Here $y = \sqrt{cx}$, the limits of integration are 0 and a,

$$P = \int_0^a 2\pi \sqrt{cx} \cdot \sqrt{1 + \left(\frac{\sqrt{c}}{2\sqrt{x}}\right)^2}\, dx$$

$$= \int_0^a 2\pi \sqrt{cx + \frac{c^2}{4}}\, dx = \int_0^a \frac{2\pi}{c} \left(cx + \frac{c^2}{4}\right)^{1/2} d\left(cx + \frac{c^2}{4}\right)$$

$$= \frac{2\pi}{c} \cdot \frac{\left(cx + \frac{c^2}{4}\right)^{3/2}}{3/2}\Bigg|_{x=a} - \frac{2\pi}{c} \cdot \frac{\left(cx + \frac{c^2}{4}\right)^{3/2}}{3/2}\Bigg|_{x=0}$$

$$= \frac{4\pi}{3c}\left(ca + \frac{c^2}{4}\right)^{3/2} - \frac{4\pi}{3c} \cdot \left(\frac{c^2}{4}\right)^{3/2}$$

$$= \frac{1}{6}\pi c^2 \left(\left(1 + \frac{4a}{c}\right)^{3/2} - 1\right).$$

6.14. The Differential Equation

A *differential equation* is a relation connecting the values of a function, its derivatives and its argument. The part of the unknown in a differential equation is played by a function which satisfies the differential equation.

We can regard the problem of integration of the function $f(x)$ as a problem of solving a differential equation $y' = f(x)$, where y is the required function.

In the simplest case the solution of a differential equation reduces to integration. This is not always possible, however, and so the problem of solving a differential equation proves to be more complicated than integration.

A differential equation of the form $y' = F(x, y)$ can be clearly represented by geometric means. This equation means that the slope of the graph of the required function (the "integral curve"), which passes through the point $M_0(x_0, y_0)$, is known beforehand, it is equal to $F(x_0, y_0)$. Thus, the equation $y' = F(x, y)$ defines the so-called *direction field*, i.e. the direction of the integral curves (i.e. tangents to them) is known at all points of the domain of definition of the function $F(x, y)$.

For instance, the equation $y' = y/x$ means that the slope of the graph of the solution, which passes through the point (x_0, y_0), is equal to y_0/x_0, i.e. to the slope of the straight line $y = \frac{y_0}{x_0} x$, which passes through the origin and the point (x_0, y_0). This means, in turn, that a tangent to the graph at any point passes through the origin (Fig. 6.29). This property is inherent in any straight line $y = cx$, which passes through the origin, since at any point the tangent coincides with the straight line itself. Indeed, the function $y = cx$ satisfies the equation $y' = y/x$ for any c.

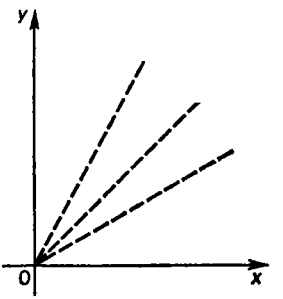


Fig. 6.29

These arguments show that the problem of solving the equation $y' = F(x, y)$ is a problem of constructing curves (the graphs of the solutions) from the given direction field. A further development of this idea leads, in particular, to the construction of efficient methods of the approximate solution of differential equations irrespective of whether the solution reduces to integration or not.

Let us consider some examples of differential equations whose solution reduces to integration.

Example 1. $y' = \frac{y}{x}$.

Solution. $\frac{y'}{y} = \frac{1}{x}$. Multiplying this relation by dx, we obtain $\frac{dy}{y} = \frac{dx}{x}$. Integrating, we find that $\ln y = \ln x + C$, whence $y = xe^C = C_1 x$.

Reasoning in this way, we can think that the constant $C_1 = e^C$ can be only positive. But this is not so. For the function $1/x$ there is an antiderivative for $x < 0$ as well, namely, $\ln(-x)$, since its derivative is equal to $\frac{1}{-x}(-x)' = \frac{1}{x}$. Thus, for any $x \neq 0$ $\ln|x|$ is an antiderivative of the function $1/x$. We infer from the equation $\frac{dy}{y} = \frac{dx}{x}$ that $\ln|y| = \ln|x| + C$, $|y| = |x| C_1$, where $C_1 = e^C >$

0, $y = \pm C_1 x$, and so in the solution $y = C_1 x$ the constant C_1 can assume positive as well as negative values. The value $C_1 = 0$ is also possible since the function $y = 0$ satisfies the equation $y' = y/x$. We have lost this solution when we divided both sides of the equation by y.

Example 2. $y' = ky$. Find a solution which is equal to $y_0 > 0$ for $x = 0$.

Solution. $\frac{y'}{y} = k$, $\frac{dy}{y} = k\,dx$, $\ln y = kx + C$, $y = e^{kx} C_1$, where $C_1 = e^C$. We have $y_0 = e^0 \cdot C_1$ for $x = 0$ whence $C_1 = y_0$. Thus $y = y_0 e^{kx}$ is the required solution.

This result can be easily interpreted. We assume that x is time and y is the mass of a substance. Let $k > 0$. In this case the equation means that the velocity of increase of the mass is proportional to the mass at every moment. This is the law of variation, for instance, of the mass of bacteria or yeast which develop in some culture medium until the properties of the medium change. Now if $k = -k_1 < 0$, then the equation $y' = -k_1 y$ shows that the rate of variation of y is negative and proportional to y, i.e. the rate of decrease of the mass of the substance is proportional to the mass. During a radioactive decay the mass of a substance varies according to the law $y = y_0 e^{-k_1 x}$, the so-called *half-life* being the convenient characteristic of the decaying substance. This is the time in which the mass of the substance decays to half its original value. The half-life T is connected with the coefficient k_1 by the relation $e^{-k_1 T} = \frac{1}{2}$, whence $T = \frac{1}{k_1} \ln 2$ and $k_1 = \frac{1}{T} \ln 2$. Substituting for k_1 its expression in terms of T, we obtain

$$y = y_0 e^{-\frac{x}{T} \ln 2} = y_0 2^{-x/T}.$$

6.15. Some Differential Equations Important for Mechanics

Newton's second law plays a fundamental part in mechanics when the motion of a particle under the action of some forces is investigated. When applied to the rectilinear motion, this law states that the *acceleration of a particle is proportional to the force applied and inversely proportional to its mass.*

With an appropriate choice of units of measurement, the acceleration a is equal to the quotient of the division of the force F by the mass m of the particle:

$$a = \frac{F}{m}$$

or, what is the same,

$$F = ma.$$

Recall that *acceleration is the rate of variation of the velocity*, i.e. is equal to the derivative of the velocity with respect to time. In turn, when the motion is rectilinear, the velocity is equal to the derivative of the coordinate of the particle with respect to time. The derivative of the derivative of the function $y = f(t)$ is the *second derivative* and is designated as y'' or $f''(t)$ (sometimes the symbol $\dot{y}$ is used for the derivative of y with respect to time and the symbol $\ddot{y}$ is used for the second derivative). Thus the acceleration of the rectilinear motion is equal to the second derivative of the coordinate y of a moving particle with respect to time. Taking this into account, we can write Newton's second law as

$$F = my''.$$

Thus, to derive an equation of motion under the action of a force F (which may depend on time, the coordinates of the particle and its velocity), we must investigate the equation

$$y'' = \frac{1}{m} F.$$

Example 1. Motion under the action of a constant force (such as, for instance, the motion by gravity at the surface of the earth).

In this case the acceleration $y'' = \frac{1}{m} F = a$ is constant. It is easy to solve the equation $y'' = a$, namely, $y' = at + c_1 = v$, $y = \frac{at^2}{2} + c_1 t + c_2$. If at the initial moment the coordinate is y_0 and the velocity is v_0, we find, by substituting $t = 0$ into the expressions for the velocity v and the coordinate y, that $c_1 = v_0$, $c_2 = y_0$. Thus the equation

of motion is

$$y = \frac{at^2}{2} + v_0 t + y_0.$$

This is a uniformly accelerated motion considered in 1.10.

Example 2. Motion under the action of a force proportional to the deviation of a particle from the equilibrium position and directed opposite to the deviation.

Assume that a particle of mass m is attached to a spring whose mass can be neglected. Assume, furthermore, that the spring becomes longer by $y = cF$ under the action of an extending force F and becomes shorter by cF under the action of a contracting force F, with the same coefficient c. The quantity $\alpha = 1/c$ is the coefficient of stiffness of the spring, and the smaller the deformation of the spring under the action of the force, the greater is this coefficient. The point that has deviated from the equilibrium position by y, is subjected to the action of the force $F = -\alpha y$. We take the minus sign since the force acts in the direction opposite to the deviation from the equilibrium position. According to Newton's second law, we obtain

$$my'' = -\alpha y, \text{ or } y'' = -\frac{\alpha}{m} y.$$

We set $\alpha/m = k^2$. The equation assumes the form $y'' = -k^2 y$. We can easily indicate some solutions of the equation, such, for instance, as $y = \cos kt$ since $(\cos kt)' = -k \sin kt$; $(\cos kt)'' = -k^2 \cos kt$, or $y = \sin kt$ since $(\sin kt)' = k \cos kt$, $(\sin kt)'' = -k^2 \sin kt$. A more general solution is

$$C_1 \cos kt + C_2 \sin kt$$

for any constants C_1 and C_2.

All solutions of the equation $y'' = -k^2 y$ are given by the formula

$$y = C_1 \cos kt + C_2 \sin kt.$$

Proof. We set $y = z \cos kt$, where z is a new unknown function. Then $y' = z' \cos kt - kz \sin kt$, $y'' = z'' \cos kt - z'k \sin kt - kz' \sin kt - k^2 z \cos kt$, whence, after the substitution into $y'' = -k^2 y$, we obtain

$$z'' \cos kt - 2kz' \sin kt - k^2 z \cos kt = -k^2 z \cos kt,$$

$$z'' \cos kt = 2kz' \sin kt.$$

Next we have

$$\frac{z''}{z'} = \frac{2k \sin kt}{\cos kt}, \quad \frac{dz'}{z'} = -\frac{2d(\cos kt)}{\cos kt}.$$

The integration yields $\ln z' = -2 \ln \cos kt + C$, $z' = \frac{c_1}{\cos^2 kt}$. We integrate once again and get

$$z = \int \frac{C_1}{\cos^2 kt}\, dt = \frac{C_1}{k} \int \frac{d(kt)}{\cos^2 kt} = \frac{C_1}{k} \tan kt + C_2,$$

$$y = z \cos kt = \frac{C_1}{k} \tan kt \cos kt + C_2 \cos kt$$

$$= \frac{C_1}{k} \sin kt + C_2 \cos kt.$$

And this is what we wished to prove since C_1/k is an arbitrary constant.

It is easy to find the constants C_1 and C_2 in the formula $y = C_1 \cos kt + C_2 \sin kt$ if the coordinate y_0 and the velocity v_0 of the particle at the moment $t = 0$ are known. Substituting $t = 0$ into the expressions for $y = C_1 \cos kt + C_2 \sin kt$ and for the velocity $v = y' = -C_1 k \sin kt + C_2 k \cos kt$, we obtain $y_0 = C_1$, $v_0 = C_2 k$ so that $C_1 = y_0$ and $C_2 = \frac{v_0}{k}$.

We can transform this solution. Setting

$$a = \sqrt{C_1^2 + C_2^2}, \quad \sin \varphi = \frac{C_1}{\sqrt{C_1^2 + C_2^2}}, \quad \cos \varphi = \frac{C_2}{\sqrt{C_1^2 + C_2^2}},$$

we obtain

$$y = a(\sin \varphi \cos kt + \cos \varphi \sin kt) = a \sin (kt + \varphi).$$

This is an *equation of harmonic oscillations with amplitude* $a = \sqrt{C_1^2 + C_2^2} = \sqrt{y_0^2 + \frac{v_0^2}{k^2}}$ *and period* $\frac{2\pi}{k}$, i. e. *with frequency* $\frac{k}{2\pi} = \frac{1}{2\pi} \sqrt{\frac{\alpha}{m}}$.

It is interesting to point out that frequency proves to be proportional to the square root of the stiffness coefficient α and inversely proportional to the square root of the mass. It is clear by intuition that the frequency must increase with stiffness and decrease with an increase in the mass. But we have got a rigorous law of variation of frequency with the variation of the stiffness and the mass.

EXERCISES TO CHAPTER 6

1. Find three antiderivatives for each of the given functions and write the general form of the antiderivative for the given function $f(x)$:

1. $f(x) = 3x^2$. 2. $f(x) = 4x^3 - 2x$.
3. $f(x) = -\sin x$. 4. $f(x) = \cos x$.

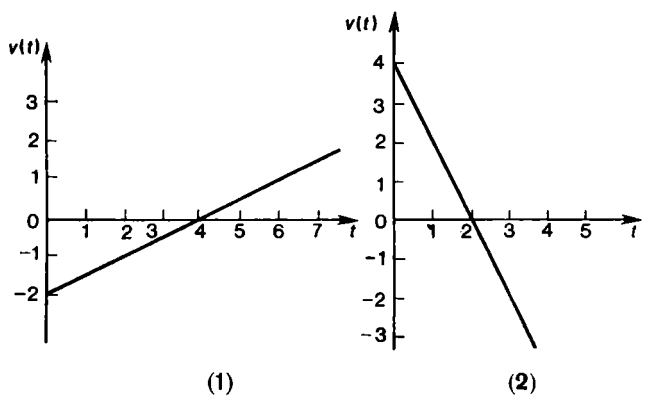
2. Find each of the following indefinite integrals:

1. $\int x^3\,dx$. 2. $\int x^{-2}\,dx$. 3. $\int \frac{1}{\sin^2 x}\,dx$.
4. $\int \frac{1}{\cos^2 x}\,dx$. 5. $\int 2\,dx$. 6. $\int dx$.

3. Find each of the following integrals and from the set of antiderivatives obtained choose the antiderivative whose graph passes through the given point $M(x, y)$:

1. $\int 4x\,dx$, $M(1, -2)$. 2. $\int \frac{1}{2}x\,dx$, $M(4, -5)$.
3. $\int \sin x\,dx$, $M\left(\frac{\pi}{2}, 3\right)$. 4. $\int \cos x\,dx$, $M\left(\frac{\pi}{6}, 1.5\right)$.

4. A point moves along a coordinate line. Proceeding from the given graph of the velocity of motion (Fig. 6.30), construct the graph



(1) (2)

Fig. 6.30

of variation of the coordinate of the point for different initial conditions.

5. Find the following integrals:

1. $\int (x-1)^3\,dx$. 2. $\int (x^2+4)(4-x^2)\,dx$.

3. $\int (2+x)(4-2x+x^2)\,dx$. 4. $\int (2+0.5x)^3\,dx$.

5. $\int \frac{3x^4-2x^2+6}{x^2}\,dx$. 6. $\int \frac{-x^4+2x-3}{x^3}\,dx$.

7. $\int \frac{3x^2-x+2}{2x\sqrt{x}}\,dx$. 8. $\int \frac{2x^2\sqrt{x}-4x\sqrt{x}+\sqrt{x}}{3x}\,dx$.

9. $\int (\sqrt{x}+\sqrt[3]{x}-\sqrt[4]{x})\,dx$. 10. $\int (\sqrt[3]{x^2}-\sqrt[5]{x^3}+\sqrt[7]{x^5})\,dx$.

6. Find the following integrals; when necessary, use the fact that $da^x = a^x \ln a\,dx$ and, hence, $\int a^x\,dx = \frac{a^x}{\ln a}+C$:

1. $\int (e^{x+1}+2^x)\,dx$. 2. $\int (3^{1+x}-e^x)\,dx$. 3 $\int (3^{2x}-3^x)\,dx$.

4. $\int \frac{2^x-3^x}{6^x}\,dx$. 5. $\int \frac{e^{2x}-1}{e^x}\,dx$.

6. $\int \left(3^x-\left(\frac{1}{3}\right)^x\right)3^x\,dx$.

7. Find the following integrals:

1. $\int \frac{x-3}{x}\,dx$. 2. $\int \frac{x^2-2x}{x^2}\,dx$.

8. Find the following integrals:

1. $\int \frac{4}{\sin^2 2x}\,dx$. 2. $\int \frac{\cos 2x}{\cos x+\sin x}\,dx$.

3. $\int \frac{\cos 2x}{\sin x-\cos x}\,dx$. 4. $\int \frac{\sin^2 x\cos x-\cos^2 x\sin x}{\sin 2x}\,dx$.

5. $\int \frac{1+\sin 2x}{\sin x+\cos x}\,dx$. 6. $\int (2+\tan^2 x+\cot^2 x)\,dx$.

7. $\int \frac{x^4-x^2-1}{x^2+1}\,dx$. 8. $\int \frac{x^3+x^2+x}{x^2+1}\,dx$.

9. Find the following integrals reducing the elements of integration to the form $f(\varphi(x))\,d\varphi(x)$:

1. $\int (1-3x)^2\,dx$. 2. $\int \frac{dx}{(0.5x-3)^2}$.

3. $\int \sqrt{\left(4+\frac{1}{3}x\right)^3}\,dx$. 4. $\int \sqrt[3]{(2+5x)^2}\,dx$.

5. $\int \cos\left(2x-\frac{\pi}{4}\right)dx$. 6. $\int \sin\left(\frac{\pi}{3}-\frac{x}{2}\right)dx$.

7. $\int \frac{dx}{1+2x}$. 8. $\int \frac{dx}{5-3x}$.

9. $\int e^{1-3x}\,dx$. 10. $\int (e^{3+2x})^{-1}\,dx$. 11. $\int 2^{-0.5x+2}\,dx$.

12. $\int 3^x \left(\frac{1}{9}\right)^{1+x} dx$. 13. $\int \frac{dx}{\sin^2(2x-1)}$.

14. $\int \frac{dx}{\cos^2\left(3-\frac{1}{2}x\right)}$. 15. $\int \sin 2x\,dx$.

16. $\int \tan x\,dx$. 17. $\int \cot x\,dx$. 18. $\int \sin^3 x\,dx$.

19. $\int \frac{x\,dx}{x^2+1}$. 20. $\int \frac{x^2\,dx}{x^2+2}$.

◀ **10.** Find the following integrals using the method of integration by parts:

1. $\int xe^{-x}\,dx$. 2. $\int x\cdot 3^x\,dx$. 3. $\int \ln^2 x\,dx$.

4. $\int \frac{\ln x}{x^3}\,dx$. 5. $\int x \sin 2x\,dx$. 6. $\int x \cos 3x\,dx$.

7. $\int e^x \sin x\,dx$. 8. $\int e^x \cos x\,dx$. 9. $\int \frac{x}{\cos^2 x}\,dx$.

10. $\int x \tan^2 x\,dx$. ▶

11. Calculate the area of the curvilinear trapezoid bounded by the graph of the function $y = f(x)$, the straight lines $x = a$, $x = b$ and $y = 0$:

1. $f(x) = 4 - x^2$, $a = -2$, $b = 2$.
2. $f(x) = x^2 + 2$, $a = -2$, $b = 2$.
3. $f(x) = \frac{-1}{x+2}$, $a = -1$, $b = 4$.
4. $f(x) = \frac{1}{x-3}$, $a = -2$, $b = 2$.
5. $f(x) = x\sqrt{4-x^2}$, $a = 0$, $b = 2$.
6. $f(x) = \frac{x}{\sqrt{4-x^2}}$, $a = 0$, $b = \sqrt{3}$.

12. Calculate the area of the figure bounded by the following curves:

1. $y = x^2 - 2x + 2$, $y = 2x - 1$.
2. $y = -x^2 - 3x + 4$, $y = -2x + 2$.
3. $y = x^2 - x - 2$, $y = -x^2 - x$.
4. $y = x - 5$, $y = -4/x$.
5. $y = \tan x$, $y = \sin x$, $x = \pi/3$.
6. $y = \cos x$, $y = \sin x$, $x = 0$.

13. Solve the following problems without resorting to the calculation of the definite integrals:

1. Using the geometric interpretation, compare $\int_0^1 dx^2\,dx$ and $\int_0^1 \sqrt{x}\,dx$ and find th sum.

2. Compare $\int_0^{\pi/2} \cos x\,dx$ and $\int_0^{\pi/2} \sin x\,dx$.

3. Calculate the area of the figure bounded by the following curves:
(a) $y = \sin x$, $y = 2$, $x = 0$, $x = 2\pi$,
(b) $y = \cos x$, $y = -1$, $x = 0$.

4. Prove the following equalities and give a geometric interpretation of them:

(a) $\int_{-\alpha}^{0} \cos x\,dx = \int_0^{\alpha} \cos x\,dx, \quad \alpha > 0,$

(b) $\int_{-\alpha}^{0} \sin x\,dx = -\int_0^{\alpha} \sin x\,dx, \quad \alpha > 0,$

(c) $\int_{-3}^{5} (x^2 - 2x + 1)\,dx = 2\int_{-3}^{1} (x^2 - 2x + 11)\,dx,$

(d) $\int_{-3}^{5} (x^3 - 3x^2 + 3x - 1)\,dx = 0.$

14. Calculate the following definite integrals:

1. $\int_{-\pi}^{\pi/2} \sin\frac{x}{2}\,dx.$ 2. $\int_{-\pi/6}^{\pi/4} \cos 2x\,dx.$

3. $\int_{-1}^{1} (3-2x)^3\,dx$. 4. $\int_{0}^{3} \frac{dx}{(1+2x)^2}$.

5. $\int_{-2}^{1} \frac{dx}{x}$. 6. $\int_{0}^{2} \frac{dx}{x-3}$. 7. $\int_{-5}^{2} |x|\,dx$.

8. $\int_{0}^{4} |1-x|\,dx$. 9. $\int_{-\pi}^{\pi} \sin^2 x\,dx$. 10. $\int_{-\pi}^{\pi} \cos^3 x\,dx$.

◀ 11. $\int_{0}^{1} xe^x\,dx$. 12. $\int_{\pi/4}^{\pi/3} \frac{x}{\sin^2 x}\,dx$.

13. $\int_{0}^{\pi/2} e^{2x} \cos x\,dx$. 14. $\int_{1}^{e} \ln^3 x\,dx$. ▶

15. Evaluate the following integrals using the relation

$$m(b-a) \leqslant \int_{a}^{b} f(x)\,dx \leqslant M(b-a),$$

where m is the least and M is the greatest value of $f(x)$ on $[a, b]$ (see the proof of the first mean value theorem):

1. $\int_{0}^{1} \frac{x}{x^3+16}\,dx$. 2. $\int_{1}^{2} \frac{x}{x^3+16}\,dx$.

3. $\int_{2}^{3} \frac{x}{x^3+16}\,dx$. 4. $\int_{3}^{4} \frac{x}{x^3+16}\,dx$.

5. $\int_{4}^{5} \frac{x}{x^3+16}\,dx$. 6. $\int_{5}^{6} \frac{x}{x^3+16}\,dx$.

16. First evaluate the integral $\int_{0}^{6} \frac{x}{x^3+16}\,dx$ using the inequality considered in Problem 15 and then get a more exact estimate using the results obtained in Problem 15.

17. Construct the graph of the function $y = \frac{x}{x^3+16}$ taking one square of your checked copybook as a scale unit laid off along the abscissa axis and 100 squares as a scale unit laid off along the axis of

ordinates. Employing the graph, find the approximate value of $\int_0^6 \frac{x}{x^3+16}\,dx$ using the following formula:

$$S \cong 0.01\left(N+\frac{n}{2}\right),$$

where N is the number of squares contained in the corresponding curvilinear trapezoid and n is the number of squares through which the graph of the function $\frac{x}{x^3+16}$ passes on the interval $[0, 6]$. Compare the answers to Problems 16 and 17.

18. Evaluate the integral $\int_0^2 \sqrt{x(x-2)(x-4)}\,dx$ using the inequality considered in Problem 15.

19. Get an estimate of the integral $\int_0^2 \sqrt{x(x-2)(x-4)}\,dx$, more accurate than in Problem 18, dividing the interval $[0, 2)$ into five (ten) smaller intervals and using the inequality to each of the intervals.

20. Let $f(x) = C$ const, $C > 0$. The graph of this function has the form shown in Fig. 6.31. The figure bounded by the curves $y = C$,

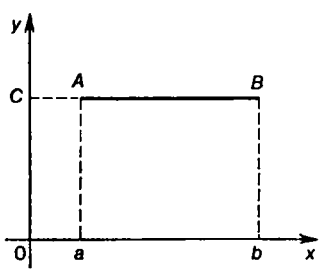


Fig. 6.31

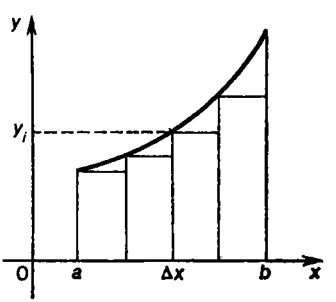


Fig. 6.32

$y = 0$, $x = a$ and $x = b$ is a rectangle whose area $S = C(b - a)$. Find the physical meaning of the area of the rectangle if

(1) the time of motion in seconds is laid off along the abscissa axis and the velocity of motion, to be more exact, the rate of variation of the path in m/s, is laid off along the axis of ordinates,

(2) the path, in metres, traversed by the point moving in the straight line is laid off along the abscissa axis and the force, in newtons, which produces the motion, is laid off along the axis of ordinates,

(3) the length of a bar, in metres, is laid off along the abscissa axis and the density of the bar, in kg/m, is laid off along the axis of ordinates,

(4) the path, in km, traversed by a car is laid off along the abscissa axis and the rate of consumption of gas, *l*/km, is laid off along the axis of ordinates,

(5) the body temperature, in degrees centigrate, is laid off along the abscissa axis and heat capacity of the body, i.e. the quantity of heat, in joules, needed to heat the body by 1°C, is laid off along the axis of ordinates.

21. The graph depicted in Fig. 6.32 shows the relationship between two physical quantities, x and y. To understand the physical meaning of the area under the graph, we divide the interval $[a, b]$ into small subintervals and assume the quantity y to be constant on each of them. We get a stepped figure whose area is approximately equal to the area of a curvilinear trapezoid. Replacing in this way the curvilinear trapezoid by a stepped figure, we make a numerical error in the estimate of the area but evidently retain the physical meaning of the area. The area of each rectangle is equal to $y_i \Delta x$ and the physical meaning of this product is defined by the measures of the quantities entering into the product, say, if the curve is the voltage of an electric field y vs time x, then $y \Delta x$ is a charge transferred in the time Δx. We can find the exact value of the area and, hence, of a physical quantity, by calculating the corresponding integral.

Find the physical meaning of the integral $\int_a^b f(x)\,dx$ and the numerical value of this physical quantity by calculating the integral, make the corresponding drawings:

1. $\int_0^{10} (2t + t^2)dt$, where t is time in seconds, $f(t) = 2t + t^2$ is the velocity of motion in m/s.

2. $\int_0^4 3x\,dx$, where x is the deformation of the spring in cm and $f(x) = 3x$ is the force in newtons needed to contract the spring by x cm.

3. $\int_4^{60} R\,(0.5 - 0.003t)^2\,dt$, where t is time in seconds and R is the resistance of the conductor in ohms, $0.5-0.003t$ is the law of variation of the current in amperes.

4. $\int_0^{100} (2 + 0.001x^2)\,dx$, where x is the length of the bar in cm and $f(x) = 2 + 0.001x^2$ is the density of the bar in g/cm.

5. $\rho \int_0^1 \frac{dl}{4 \cdot 10^{-6} + 10^{-4} l^2}$, where l is the length of the conductor in metres, ρ is the resistivity of the cross-section of the conductor in $\frac{\Omega \cdot m^2}{m}$, $\frac{1}{4 \cdot 10^{-6} + 10^{-4} l^2}$ is the cross-sectional area of the conductor in square metres dependent on the distance of the cross-section from the end of the conductor.

6. $\frac{1}{S} \int_0^1 (2 \cdot 10^{-8} + 10^{-9} l)\, dl$, where l is the length of the conductor in metres, S is the cross-sectional area in square metres, $\rho(l) = 2 \cdot 10^{-8} + 10^{-9} l$ is the resistivity of the conductor in $\frac{\Omega \cdot m^2}{m}$ at the distance l from the end of the conductor.

7. $\int_0^{10} \pi (2 + 0.03l)^2\, dl$, where l is the length of the bar in cm, $R(l) = 2 + 0.03l$ is the radius of the cross-section of the bar at the distance l from one of its ends, the cross-sectional area is a circle.

22. Assume that $f(x)$ is a function continuous on $[a, b]$ and the point $x = c$ is such that $\int_a^b f(x)\, dx = f(c)(b - a)$. Then $f(c) = \frac{1}{b-a} \int_a^b f(x)\, dx$ is the *mean value of the function on the interval* $[a, b]$.

In problems with real data, the mean value of a function has a clear physical meaning. In the following examples, we must find the mean value of a function on a given interval and find its physical meaning proceeding from the conditions of the problem:

1. $v(t) = 2 + 3t$ is the velocity of motion in m/s, t is time in seconds, $t \in [2, 8]$.
2. $v(t) = t^2$ is the velocity of motion in m/s, t is time in seconds, $t \in [0, 10]$.
3. $F(s) = 40 + 4s$ is a force in newtons, s is the path in metres traversed under the action of this force, $s \in [0, 20]$.
4. $a(t) = \sqrt{t}$ is the acceleration of motion in m/s^2, t is time in seconds, $t \in [0, 100]$.
5. $S(x) = 0.01\, \pi x^2$ is the cross-sectional area of the bar in cm^2 at the distance x from one of its ends, $x \in [0, 10]$, x is in cm and the cross-section is a circle.

23. Use a sheet of checked paper to construct a broken line $ABCD$: $A(2, 4)$, $B(6, 10)$, $C(8, 6)$, $D(13, 8)$. Assuming the broken line to be the graph of the function $y = f(x)$, find the mean value of $f(x)$ on the interval $[2, 13]$.

24. Calculate the upper and the lower integral sum for the integral $\int_0^1 x^2\, dx$, assuming that the interval $[0, 1]$ is divided into n equal parts.

Find the limits of these sums and compare these limits with the value of the integral.

25. Calculate the upper and the lower integral sum for the integral $\int_0^{\pi} \sin x\, dx$ dividing the interval $[0, \pi]$ into 12 equal parts. Compare the sums with the exact value of the integral. Repeat the actions dividing the interval $[0, \pi]$ into 24 equal parts. When seeking the solution, use the symmetry of the part of the sine curve with respect to the straight line $x = \pi/2$.

26. Construct the graph of the function $F(x)$:

1. $F(x) = 1 + \int_0^x 2t\, dt$, $x \in [0, 4]$.

2. $F(x) = \int_0^x \cos t\, dt - 1$, $x \in [0, 2\pi]$.

27. On a given interval construct a broken line which can approximately replace the graph of the function $F(x)$ given as an integral with variable upper limit; first set up an integral sum which would express the approximate value of the integral:

1. $F(x) = \int_0^x \frac{-t^2 - 4t + 20}{(t+2)^2}\, dt$, $x \in [0, 10]$. Divide the interval $[0, 10]$ into ten equal parts: $[0, 1]$, $[1, 2]$ and so on. When setting up the integral sum, take the value of the integrand function at the midpoints of the subintervals, i.e. at the points $z_1 = 0.5$, $z_2 = 1.5$ and so on.

2. $F(x) = -1.6 + \int_5^x \frac{t^2 - 10t + 9}{(t-5)^2}\, dt$, $x \in [-5, 3]$.

Divide the interval $[-5, 3]$ into ten equal parts and take the values of the integrand function at the midpoints of the subintervals.

28. Use the rule of rectangles to make an approximate calculation of the integral. We can estimate the error of calculation proceeding from the fact that with an increase in n (the number of divisions of the integration interval), the error decreases in much the same manner as $1/n^2$, i.e. when n increases two times, the error decreases about four times. Assume that S_n and S_{2n} are approximate values of the integral which correspond to the division of the integration interval into n and $2n$ subintervals, and then $|S_n - S_{2n}|$ is approximately equal to the trebled error of approximation S_{2n}. The final estimation of the error α upon the division of the integration interval into $2n$ parts can be made with the use of the formula $\alpha \cong \frac{1}{3} |S_{2n} - S_n|$.

1. $\int_0^1 \sqrt{2-4x+4x^2}\,dx$, (a) $n=5$, (b) $n=10$.

2. $\int_0^1 \frac{dx}{1+x} = \ln 2$, (a) $n=5$, (b) $n=10$.

3. $\int_0^2 \frac{dx}{\sqrt{1+x^2}}$, (a) $n=5$, (b) $n=10$.

29. Make an approximate calculation of the integrals considered in Problem 23 using the trapezoidal rule. The error of calculation made with the use of the trapezoidal rule is estimated in the same way as that made according to the rule of rectangles.

30. Make an approximate calculation of the integrals considered in Problem 28 using Simpson's rule. With an increase in n, the error of calculation carried out in accordance with Simpson's rule decreases in much the same manner as $1/n^4$, i.e. when n increases two times, the error of calculation decreases approximately 16 times, and we can therefore estimate the error employing the formula

$$\alpha \cong \frac{1}{15}\,|S_{2n}-S_n|.$$

31. Calculate the volume of a solid of revolution resulting from the rotation, about the x-axis, of a curvilinear trapezoid bounded by the graph of the function $y=f(x)$, the straight lines $x=a$ and $x=b$ and the x-axis:

1. $f(x)=2+2x$, $a=0$, $b=4$.
2. $f(x)=x^2+1$, $a=0$, $b=3$.
3. $f(x)=xe^x$, $a=0$, $b=1$.
4. $f(x)=\frac{1}{2}(e^x+e^{-x})$, $a=-1$, $b=1$.

32. Calculate the volume of a solid resulting from the rotation of a figure bounded by the curves $y=x^2$ and $y^2=x$ about the x-axis.

33. Calculate the volume of a solid resulting from the rotation of a figure bounded by the parabola $y=2x-x^3$ and the abscissa axis about the axis of ordinates.

34. Calculate the volume of a solid resulting from the rotation of a figure bounded by the curves $y=x^2-4$, $y=0$, $x=0$ and $x=3$ about the abscissa axis.

35. Calculate the length of a curve defined by the equation $\sqrt[3]{x^2}+\sqrt[3]{y^2}=4$. A curve of this kind is called an *astroid* and its approximate shape is given in Fig. 6.33.

Hint. It pays to recall how the derivative of a function defined implicitly by an equation can be found.

36. Make an approximate calculation of the length of the ellipse $x^2 + \frac{1}{4} y^2 = 1$.

37. Make an approximate calculation of the length of a part of the hyperbola $y = 1/x$ included between the points with abscissas $x = 0.5$ and $x = 2$.

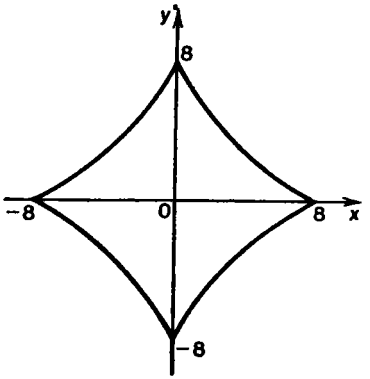


Fig. 6.33

38. Make an approximate calculation of the length of the arc of the sine curve $y = \sin x$ for $x \in [0, \pi]$.

39. Make an approximate calculation of the length of the arc of the parabola $y = x^2$ for $x \in [-2, 2]$.

40. Find the lateral area of a cone with altitude h and the radius of the base R, representing its surface as the result of the rotation of a segment of a straight line, passing through the origin, about the abscissa axis.

41. Find the surface area of a spherical band with altitude h if the radius of the corresponding sphere is r.

42. Calculate the area of the surface resulting from the rotation of a part of the cubic parabola $y^2 = \frac{1}{3} x^3$, included between the points with abscissas $x_1 = 0$ and $x_2 = a$, about the abscissa axis.

43. Make an approximate calculation of the area of the surface resulting from the rotation of one arc of the sine curve $y = \sin x$, $x \in [0, \pi]$, about the abscissa axis.

44. Find the area of the surface resulting from the rotation of a loop of the semicubic parabola $y^2 = \frac{1}{9} x (3-x)^2$ about the abscissa axis.

45. Use graphical means to solve the differential equation $y' = xy^2 + 1$ on the interval $[0, 1]$ for the initial conditions $y(0) = 0$.

Solution. We divide the interval $[0, 1]$ into four equal parts: $[0, 0.25]$, $[0.25, 0.5]$, $[0.5, 0.75]$, $[0.75, 1]$ and on each subinterval replace the integral (required) curve by a segment of the tangent to that curve at the left endpoint of the subinterval. Thus the curve passes through the point $(0, 0)$ (by the hypothesis). We calculate the derivative $y'_{x=0} = y'(0) = 0 \cdot 0 + 1 = 1$, which is the slope of the tangent to the required curve at the point $(0, 0)$. The tangent has the form $y = 1 \cdot x$. We assume that on the subinterval $[0, 0.25]$ the tangent $y = x$ is a sufficiently good substitution for the curve and, consequently, $y(0.25) = 0.25$. We seek $y'_{x=0.25} = 0.25 \cdot 0.25^2 + 1 = 1.016$, which is the slope of the tangent at the point $(0.25, 0.25)$. The equation of the tangent is $y = 0.25 + 1.016(x - 0.25)$. We replace the curve on the subinterval $[0.25, 0.5]$ by the segment of this tangent, $y(0.5) = 0.25 + 1.016 \cdot 0.25 = 0.504$. Acting by analogy, on the sub-

interval [0.5, 0.75] we get $y'(0.5) = 0.5 \cdot 0.504^2 + 1 = 1.127$, $y = 0.504 + 1.127\,(x - 0.5)$, $y(0.75) = 0.504 + 1.127 \cdot 0.25 = 0.786$. On the subinterval [0.75, 1] we get $y'(0.75) = 0.75 \cdot 0.786^2 + 1 = 1.463$, $y = 0.786 + 1.463\,(x - 0.75)$, $y(1) = 0.786 + 1.463 \cdot 0.25 = 1.152$. We tabulate the results:

x	0	0.25	0.5	0.75	1
y	0	0.25	0.504	0.786	1.152

The broken line passing through the tabulated points is an approximate graph of the required function. It is clear that by increasing the number of divisions of the interval [0, 1], we can get a more and more accurate representation of the graph of the function which is the solution of this differential equation for the given initial conditions.

46. Use graphical means to solve the equation $y' = \frac{y}{x} + 0.5x$ on the interval [4, 6] if $y(4) = -4$. Divide the interval [4, 6] into ten parts.

47. Use graphical means to solve the equation $y' = 0.1\,(x^2 + y^2)$ on the interval [1, 5] if $y(1) = 1$.

48. Use graphical means to solve the equation $y' = \frac{1}{x^2 + y^2}$ on the interval [0.5, 3.5] if $y(0.5) = 0.5$.

49. Use the method of substitution to verify that the function $y = Cx^3$ is a solution of the differential equation $3y - xy' = 0$. Construct the graphs of the solutions passing through the points

(1) (1, 1/3), (2) (1, 1), (3) (1, −1/3).

50. Use the method of substitution to verify that the function $y = C \cdot e^{2x}$ is a solution of the differential equation $y' = 2y$. Construct the graphs of the solutions passing through the points

(1) (0, 1), (2) (0, 1/5).

51. Use the method of substitution to verify that each of the functions (1) $y = \sin 2x$, (2) $y = \cos 2x$, (3) $y = \sin 2x + \cos 2x$, (4) $y = 2 \sin 2x + 3 \cos 2x$ is a solution of the equation $y'' = -4y$. Show that any function of the form $y = C_1 \sin 2x + C_2 \cos 2x$, where C_1 and C_2 are constants, is a solution of this differential equation.

52. A force varying according to the law $F(t) = \cos t$ acts on a particle of mass m. Write the law of motion of the particle (the law of variation of the coordinate of the particle) in the differential form, solve this equation in the general form and find its particular solution satisfying the conditions $x(0) = 0$ and $v(0) = 0$, where x is the coordinate of the particle and v is the velocity of motion of the particle. Construct the graph of the motion assuming that $m = 1$.

53. Solve a problem similar to Problem 52 but for a force varying according to the law $F(t) = \sin t$.

54. Write the general solution of the differential equation of rectilinear motion of a 1 kg particle under the action of a constant force of 10 N. Find particular solutions of this equation satisfying the given

initial conditions and construct the graphs of the motions:

1. $x(0) = 0,\ v(0) = 0.$ 2. $x(0) = -5,\ v(0) = -4.$
3. $x(0) = 10,\ v(0) = -20.$ 4. $x(0) = -4,\ v(0) = 2.$

55. Write the general solution of the differential equation of harmonic oscillations. Find its particular solution satisfying the given initial conditions and write it in the form

$$x(t) = A\cos(\omega t + \varphi), \text{ where } A > 0,\ \varphi \in [0, 2\pi]:$$

1. $x'' = -25x,\quad x(0) = 1,\ x'(0) = 5.$
2. $x'' = -4x,\quad x(0) = 1,\ x'(0) = -2.$
3. $x'' = x,\quad x(0) = 0,\ x'(0) = \sqrt{3}.$

56. Write the general solution of the differential equation. Find its particular solution satisfying the given initial conditions:

1. $y' = 3y,\ y(0) = 2,$ 2. $y' = -2y,\ y(1) = e.$

57. Calculate the fraction of radium which will remain in 1000 years if its half-life is 1550 years.

58. The half-life of a radioactive substance is one hour. In how many hours will its quantity be equal to 1/10 of its initial value?

59. A point moves along a coordinate line according to the law $x = f(t)$, where x is the coordinate of the point on the coordinate line in metres and t is time in seconds. Show that the path traversed by the point during the time from t_1 to t_2 is

$$S = \int_{t_1}^{t_2} |v(t)|\, dt,$$

where $v(t) = \dfrac{dx}{dt}$. Find the path traversed by the point moving according to the given law during the indicated time:

1. $x(t) = t^2 - 4t + 2,\quad t \in [0, 6].$
2. $x(t) = \dfrac{1}{3}t^3 - 3t^2 + 8t - 4,\quad t \in [0, 10].$

60. Find the volume of a spherical layer (a part of the sphere lying between parallel planes) if the distances from the sections bounding it to the centre of the sphere are a and b respectively and the radius of the sphere is R. Obtain a particular form of the formula if one of the distances is equal to the radius of the sphere.

61. A sphere of radius R submerged in water by $\dfrac{1}{3}R$ floats on its surface. Find the work that must be done to submerge the sphere entirely in water.

62. A cylinder of radius R and altitude H, submerged in water by $\dfrac{2}{3}H$, floats in it in the vertical position. What work must be done to take the cylinder out of the water?

63. Find the pressure exerted on the wall of an aquarium filled with water if the wall is shaped as a rectangle with altitude h and width d.

64. Find the force of gravitational interaction between a particle of mass m and a homogeneous bar of length l and mass M lying on the same straight line. The distance from the particle to the nearest end of the bar is c.

65. Find the kinetic energy of a homogeneous bar of length l and mass m rotating in a horizontal plane with constant angular velocity ω about a vertical axis which passes through one of the ends of the bar.

◀ Additional Problems

66. Find the derivative of the function $F(x)$:

1. $F(x)=\int\limits_1^x \sin t\,dt$.
2. $F(x)=\int\limits_x^1 \ln t\,dt$, $x>0$.
3. $F(x)=\int\limits_0^{x^2} \frac{dt}{\sqrt{1+t^4}}$.
4. $F(x)=\int\limits_x^{x^2} \sqrt{(1-t)^3}\,dt$.

67. Find the greatest value of the function $F(x)$ on a given interval:

1. $F(x)=\int\limits_0^x (2t-t^2)\,dt$, $x\in[0,\ 4]$.
2. $F(x)=\int\limits_x^5 (t^2+3t)\,dt$, $x\in[-3,\ 5]$.
3. $F(x)=\int\limits_4^x (t^2-8t+16)\,dt$, $x\in[0,\ 5]$.
4. $F(x)=\int\limits_x^\pi \cos t\,dt$, $x\in\left[\frac{\pi}{2},\ \frac{3}{2}\pi\right]$.

68. Find the least value of the function $F(x)$ on a given interval

1. $F(x)=\int\limits_0^x \frac{2t+1}{t^2-2t+2}\,dt$, $x\in[-1,\ 1]$.
2. $F(x)=\int\limits_x^2 \log_{1/3} t\,dt$, $x\in\left[\frac{1}{10},\ 4\right]$.

69. Find the points of extremum and the intervals of monotonicity of each of the given functions:

1. $F(x)=\int_0^x (t-1)(t-2)^2\,dt$.

2. $F(x)=\int_0^x \frac{(t^2-3t+4)(t^2-9)}{t^2+t+1}\,dt$.

70. Find the limit of each of the given functions and give a geometric interpretation of the limit:

1. $\lim\limits_{x\to 0}\int_x^1 \frac{dt}{\sqrt{t}}$. 2. $\lim\limits_{x\to 0}\int_x^1 \frac{dt}{t}$.

3. $\lim\limits_{x\to\infty}\int_0^{\infty} e^{-at}\,dt, \quad a>0$.

71. Prove that if $f(x)$ is an odd function integrable on the interval $[-a, a]$, then $\int_{-a}^{a} f(x)\,dx = 0$.

72. Prove that if $f(x)$ is an even function integrable on the interval $[-a, a]$, then

$$\int_{-a}^{a} f(x)\,dx = 2\int_0^a f(x)\,dx.$$

73. Let m and n be natural numbers. Prove that

1. $\int_0^{2\pi} \cos mx\cdot\sin nx\,dx=0$,

2. $\int_0^{2\pi} \sin mx\cdot\sin nx\,dx=\begin{cases} 0, & \text{if } m\neq n,\\ \pi, & \text{if } m=n,\end{cases}$

3. $\int_0^{2\pi} \cos mx\cdot\cos nx=\begin{cases} 0, & \text{if } m\neq n,\\ \pi, & \text{if } m=n.\end{cases}$

74. Show that the solution of the differential equation $y' = f(x)$ with the initial condition $y(x_0) = y_0$ can be written as

$$y = y_0 + \int_{x_0}^{x} f(t)\,dt$$

and vice versa.

75. Apply Picard's method to find approximate solutions of the equation $y' = x + y^2$, $y = 0$ for $x = 0$. Here $f(x, y) = x + y^2$.

We shall first consider the method in the general form. We have a differential equation $y' = f(x, y)$, where y is an unknown function of x whose approximate value at the point x must be found under the condition that $y = y_0$ for $x = x_0$. We take $y(x) = y_0$ as a *zero* (initial), very rough, approximation. We get the *first* approximation replacing y in $f(x, y)$ by y_0. We obtain

$$y_1' = f(x, y_0).$$

Integrating with due account of the initial conditions, we get

$$y_1 = y_0 + \int_{x_0}^{x} f(t, y_0)\,dt.$$

We get the *second* approximation by replacing y in $f(x, y)$ by y_1. We have

$$y_2' = f(x, y_1)$$

and, after integration, we obtain

$$y_2 = y_0 + \int_{x_0}^{x} f(t, y_1)\,dt.$$

Replacing y in $f(x, y)$ by the second approximation and integrating, we get the *third* approximation and so on.

Thus, according to Picard's method, the successive approximations to the solution of the differential equation $y' = f(x, y)$ with the initial condition $y = y_0$ for $x = x_0$ can be obtained from the following recursion formulas:

$$y_0(x) = y_0,$$

$$y_i(x) = y_0 + \int_{x_0}^{x} f(t, y_{i-1}(t))\,dt, \quad i = 1, 2, \ldots .$$

To obtain numerical results when solving differential equations by the indicated method, we must resort to calculators.

Example. $y' = x + y^2$, $y = 0$ for $x = 0$. Here $f(x, y) = x + y^2$. The *zero* (initial) approximation is $y_0(x) = 0$.

The *first* approximation: in the expression $x + y^2$ we set $y = 0$ and this yields $y_1' = x$ and

$$y_1 = \int_0^x t\,dt = \frac{1}{2}x^2.$$

The *second* approximation: in the expression $x + y^2$ we set $y = \frac{1}{2}x^2$ and this yields $y_2' = x + (0.5x^2)^2$ and

$$y_2 = \int_0^x (t + 0.25t^4)\,dt = \frac{1}{2}x^2 + \frac{1}{20}x^5.$$

The *third* approximation: in the expression $x + y^2$ we set $y = \frac{1}{2}x^2 + \frac{1}{20}x^5$ and this yields

$$y_3' = x + \left(\frac{1}{2}x^2 + \frac{1}{20}x^5\right)^2,$$

$$y_3 = \int_0^x \left(t + \frac{1}{4}t^4 + \frac{1}{20}t^7 + \frac{1}{400}t^{10}\right) dt$$

$$= \frac{1}{2}x^2 + \frac{1}{20}x^5 + \frac{1}{160}x^8 + \frac{1}{4400}x^{11}.$$

Continuing this process, we can obtain the fourth, the fifth, and subsequent approximations to the value of the function at the point x. When definite conditions are satisfied, every new approximation gives a more accurate value of the function at the point x and, on the other hand, for x sufficiently close to x_0 and when a very high accuracy is not required, the second and the third approximations yield satisfactory results. The difference of the $(i + 1)$th and the ith approximation gives a rough idea about the accuracy of the ith approximation. Thus, in the example we have considered, for $x = 0.3$ the first approximation gives $y_1 = \frac{1}{2}(0.3)^2 = 0.045$, the second approximation gives $y_2 = 0.045 + 0.0001215 = 0.0451215$, the third approximation gives $y_3 = 0.0451215 + 0.00000041 = 0.045121191$. The third approximation adds 0.00000041 to the second and the successive approximations make a still smaller correction. Therefore we can assume that the second approximation gives the value of the function at the point $x = 0.3$ with an accuracy to within 0.000001, i.e. $y\,(0.3) \cong 0.45122$.

76. Obtain the first three approximations to the solution of each of the following differential equations using Picard's method:

1. $y' = 2y - 2x^2 - 3$, $y = 2$ for $x = 0$.
2. $y' = 2 - \frac{y}{x}$, $y = 2$ for $x = 1$.

3. $y' = xy^3 - 1$, $y = 0$ for $x = 0$; calculate the approximate value of the function at the point $x = 1$. Restrict your calculations to 0.01.

77. Construct the graphs of successive approximations y_0, y_1, y_2, y_3 to the solution of the differential equation $y' = 2 - \frac{y}{x}$ on the interval [0.8, 1.2] if $y(1) = 2$. To construct the graphs, fill in the following table:

y_i	x				
	0.8	0.9	1	1.1	1.2
y_0					
y_1					
y_2					
y_3					

78. 1. Construct the graph of the fourth approximation to the solution of the equation $y' = x - y$, obtained by Picard's method, on the interval [0.7, 1.3] if $y(0) = 1$.

2. Verify that the function $y(x) = 2e^{-x} + x - 1$ is a solution of the equation $y' = x - y$. Compare the values of the exact solution with that of the fourth approximation to the solution at the points 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3.

79. Construct the graph of the third approximation to the solution of the equation $y' = 2x - \frac{y}{x}$, obtained by Picard's method, on the interval [0.7, 1.3] if $y(1) = 2$.

2. Compare the approximate solution with the exact one which is given by the formula $y(x) = \frac{2}{3}x^2 + \frac{4}{3x}$. ▶

Chapter 7

COMPLEX NUMBERS

7.1. Introductory

Students know by now that there are no square roots of negative numbers among real numbers. However, algebra itself and its applications require such an extension of the concept of a number for which the extraction of the square root of a negative number would be possible.

We have repeatedly encountered the notion of extension of a number.

Fractional numbers were introduced to make it possible to divide one integral number by another, negative numbers were introduced to make it possible to subtract a larger number from a smaller one, and irrational numbers became necessary in order to describe the result of measurement of the length of a segment in the case when the segment is incommensurable with the chosen unit of length. The addition of every succeeding class of numbers to the preceding class extends the concept of a number and the sphere of application of this concept.

The square root of the number -1 is usually denoted by the letter i and numbers of the form $a + bi$, where a and b are ordinary real numbers, are known as *complex numbers.*

However, a long time passed before complex numbers gained recognition as a valuable means of investigation. The symbol i itself (imaginary unit number), which obviously had no sense among real numbers, caused distrust.

The necessity of considering complex numbers first arose in the sixteenth century when several Italian mathematicians discovered the possibility of algebraic solution of third-degree equations.

Here is an example showing how this discovery occurred.

Assume that we have to solve an equation

$$x^3 - 6x + 4 = 0,$$

We set $x = y + \frac{2}{y}$, where y is a new unknown. As a result of substitution, we obtain an equation

$$y^3 + 6y + \frac{12}{y} + \frac{8}{y^3} - 6y - \frac{12}{y} + 4 = 0.$$

Collecting terms, we get

$$y^3 + \frac{8}{y^3} + 4 = 0.$$

Setting $y^3 = z$, we get a new equation $z + \frac{8}{z} + 4 = 0$, which is equivalent to the quadratic equation

$$z^2 + 4z + 8 = 0,$$

whence it follows that

$$z = -2 \pm \sqrt{4-8} = -2 \pm \sqrt{-4}$$
$$= -2 \pm 2\sqrt{-1} = -2 \pm 2i.$$

We have arrived at a meaningless result if we remain in the domain of real numbers. A conclusion that suggests itself is that the initial equation has no real roots. But this is not so. The root $x = 2$ arrests attention.

Let us continue our reasoning. We consider the root $z = -2 + 2i$ and write it in the following form:

$$-2 + 2i = 1 + 3i - 3 - i = 1 + 3i + 3i^2 + i^2 \cdot i$$
$$= 1 + 3i + 3i^2 + i^3 = (1 + i)^3.$$

Thus, in this case,

$$y^3 = (1 + i)^3.$$

It is natural to assume that $y = 1 + i$, and then

$$x = 1 + i + \frac{2}{1+i} = 1 + i + \frac{1-i^2}{1+i} = 1 + i + 1 - i = 2.$$

Note that taking another value of z, namely, $z = -2 - 2i$, we get nothing new, to wit,

$$-2 - 2i = (1 - i)^3,$$

whence $y = 1 - i$, and $x = 1 - i + \frac{2}{1-i} = 1 - i + 1 + i = 2$.

However, we have not got all the roots of the equation. We know from algebra that the left-hand side $x^3 - 6x + 4$

of the initial equation is divisible by $x - 2$ since it vanishes for $x = 2$. Carrying out the division, we obtain

$$(x^3 - 6x + 4) = (x - 2)(x^2 + 2x - 2).$$

Equating the first factor to zero, we get the first root, $x = 2$. Equating the second factor to zero, we obtain

$$x = -1 \pm \sqrt{3}.$$

A natural question arises: where are these two roots? The answer may seem to be unexpected. It turns out that when we extracted the cubic root, we did not take into account all possible values. The extraction of a cubic root of $(1 + i)^3$ is equivalent to the solution of the equation $y^3 - (1 + i)^3 = 0$. Factoring the left-hand side, we obtain

$$(y - 1 - i)(y^2 + (1 + i)y + (1 + i)^2) = 0.$$

Equating the first factor to zero, we get $y = 1 + i$ and equating the second factor to zero, we get

$$y = \frac{-1 - i \pm \sqrt{(1+i)^2 - 4(1+i)^2}}{2} = (1+i)\frac{-1 \pm i\sqrt{3}}{2}.$$

Substituting these two values of y into the equation $x = y + \frac{2}{y}$, we get the required two roots. We omit the corresponding calculations.

This method is of a general character. It can be applied to any equation of the form $x^3 + px + q = 0$ (the necessary substitution is $x = y - \frac{p}{3y}$) irrespective of whether it has an obvious root or not. But it turns out that if an equation including real coefficients has three distinct real roots, i.e. if the graph of the left-hand side cuts the abscissa axis at three points, then complex numbers from which cubic roots must be extracted inevitably spring up in the process of solution of the equation by this method. At the end, the terms containing the number i cancel each other and we get real answers. It has been proved that if we use some other algebraic method to solve a cubic equation with real coefficients which has three real roots, we still cannot do without extracting a cubic root of a complex number. (We understand the algebraic solution to be the representation of the

roots of an equation in terms of its coefficients as a result of arithmetic operations and the operation of the extraction of a root.)

It should also be pointed out that when we extract a cubic root in the domain of complex numbers, we get three values rather than one as we do in the domain of real numbers. To construct all the roots of the equation $x^3 + px + q = 0$, we have to take into account all three values of the cubic root.

When we pass to the domain of complex numbers, many propositions from algebra become simple and symmetric. Thus a quadratic equation with real or complex coefficients always has two roots, distinct or coincident. An nth-degree root of a nonzero real or complex number always has n different values, etc.

At the same time, the theoretical and applied value of complex numbers is far beyond the scope of algebra. The theory of functions of a complex variable which was much advanced in the nineteenth century, proved to be a very valuable apparatus for the investigation of almost all the divisions of theoretical physics, such, for instance, as the theory of oscillations, hydrodynamics, the theory of elementary particles, etc.

7.2. Basic Definitions

Complex numbers are expressions (pictures, if you like) of the form $a + bi$, where a and b are real numbers and i is a formal symbol (a letter). In this case we assume that

(1) $a + bi = c + di$ if and only if $a = c$ and $b = d$,

(2) addition is defined by the law

$$(a + bi) + (c + di) = a + c + (b + d)\, i,$$

(3) multiplication is defined by the law

$$(a + bi)(c + di) = ac - bd + (ad + bc)\, i,$$

(4) $a + 0i = a$.

The first law shows that two complex numbers are assumed to be equal if their notations are similar.

The second law signifies that addition is carried out according to the ordinary rule of addition of polynomials with collection of like terms.

The third law means that multiplication of complex numbers is carried out according to the ordinary rule of multiplication of polynomials with the only difference that the product $i \cdot i$ is replaced by -1.

The fourth law makes it possible to consider real numbers to be a special case of complex numbers when the coefficient of i is zero.

The real numbers a and b of which the complex number $a + bi$ is composed are the *components* of this number. The first component a is the *real part* of the number $a + bi$ and the second component b is the *imaginary part* (note that the imaginary part of a complex number is a real number). The fourth law shows that a complex number with the imaginary part equal to zero is a real number. A complex number with the real part equal to zero is known as a *pure imaginary number*. The square of a pure imaginary number, i.e. the product of that number by itself, is equal to a negative real number, i.e. $(bi)^2 = bi \cdot bi = -b^2$.

These four laws which enter into the *definition* of a complex number are actually connected with real numbers, the components of a complex number. They can be presented without resort to the number i. It is sufficient to write a pair of real numbers (a, b) instead of $a + bi$ and then the four laws acquire the following form:

1. $(a, b) = (c, d)$ if and only if $a = c$ and $b = d$.
2. $(a, b) + (c, d) = (a + c, b + d)$.
3. $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$.
4. $(a, 0) = a$.

In this form of notation of complex numbers and operations involving them no doubts can arise as to the consistency of this concept, we simply speak of pairs of real numbers which are acted upon according to laws 1-4.

It is easy to pass from the notation of a complex number as a pair of numbers to the ordinary notation, namely, according to laws 1-4

$$(a, b) = (a, 0) + (0, b) = (a, 0) + (b, 0)\,(0, 1)$$
$$= a + b\,(0, 1).$$

We denote the pair $(0, 1)$ by the letter i. We obtain $(a, b) = a + bi$ with $i^2 = (0, 1)\,(0, 1) = (-1, 0) = -1$.

Operations involving complex numbers possess definite properties which are the same as those of similar operations involving real numbers, to wit, addition and multiplication possess the properties of commutativity and associativity, and addition and multiplication are connected by the property of distributivity. All this can be easily derived from laws 1-4. We omit the verification of these properties.

The product of any complex number $a + bi$ by zero $0 = 0 + 0i$ is evidently zero. According to law 3, multiplication of a complex number $a + bi$ by a real number $m = m + 0i$ is equal to $ma + mbi$, i.e. multiplication of a complex number by a real number is carried out component by component. It follows immediately that the division by the nonzero real number m is also carried out component by component since the division by m is equivalent to the multiplication by $1/m$.

The complex numbers $a + bi$ and $a - bi$ which differ by the sign of the imaginary part are called *conjugate complex numbers*. If $a + bi \neq 0$, then

$$(a + bi)(a - bi) = a^2 + b^2$$

is a positive real number.

We shall show that for any complex number α there is an opposite number $-\alpha$, i.e. a number such that $\alpha + (-\alpha) = 0$. We assume that $\alpha = a + bi$ and add it to the number $-a - bi$. According to laws 2 and 4 we have

$$\begin{aligned}(a + bi) + (-a - bi) &= (a - a) + (b - b)\, i \\ &= 0 + 0i = 0,\end{aligned}$$

and so $-a - bi = (-\alpha)$. Note that $-\alpha$ can be interpreted as $(-1)\cdot\alpha$. Indeed, according to the multiplication rule, we have $(-1)\cdot(a + bi) = -a - bi$.

We shall verify whether the equation $\alpha + x = \beta$ has a unique solution. We shall first show that $x = (-\alpha) + \beta$ is a solution. We have

$$\alpha + ((-\alpha) + \beta) = \alpha + (-\alpha) + \beta = \beta,$$

and consequently $(-\alpha) + \beta$ satisfies the equation. Let us make sure that there are no other solutions. We assume that x is a solution of the equation $\alpha + x = \beta$, and then $(-\alpha) + \alpha + x = \beta + (-\alpha)$, whence it follows that $x = \beta + (-\alpha)$, and this is what we wished to prove.

The number $\beta + (-\alpha)$, which is a solution of the equation $\alpha + x = \beta$, is the *difference* of the numbers β and α and is designated as $\beta - \alpha$. Thus, by definition,

$$\beta - \alpha = \beta + (-\alpha).$$

Let us show that subtraction of complex numbers, as well as addition, is carried out component by component. Let

$$\beta = (c + di), \quad \alpha = a + bi$$

and then $\beta - \alpha = (c + di) + (-a - bi) = (c - a) + (d - b)\, i$.

It is easy to see that for any nonzero complex number $\alpha = a + bi$ there is an inverse number α^{-1}, i.e. a number such that $\alpha \cdot \alpha^{-1} = 1$. We can see that the number $(a - bi)\frac{1}{a^2+b^2}$ is the inverse of $\alpha = a + bi$. Indeed,

$$(a + bi)(a - bi)\frac{1}{a^2+b^2} = (a^2 + b^2)\frac{1}{a^2+b^2} = 1.$$

It is now easy to verify that for $\alpha \neq 0$ the equation $ax = \beta$ has a unique solution. Indeed, $\alpha(\alpha^{-1}\beta) = \alpha\alpha^{-1}\beta = \beta$, and so $x = \alpha^{-1}\beta$ is a solution of the equation $\alpha x = \beta$. In turn, if $\alpha x = \beta$, then $\alpha^{-1}\alpha x = \alpha^{-1}\beta$ and $x = \alpha^{-1}\beta$. Thus every solution of the equation $\alpha x = \beta$ coincides with $x = \alpha^{-1}\beta$ and so the solution is unique. The solution of the equation $\alpha x = \beta$ is naturally a *quotient* of the division of β by α and is designated as β/α. Let $\beta = c + di$, $\alpha = a + bi$. Then $\alpha^{-1} = \frac{a-bi}{a^2+b^2}$ and $\frac{\beta}{\alpha} = \alpha^{-1}\beta = \frac{(a-bi)(c+di)}{a^2+b^2}$. Instead of memorising this formula, you must remember that the result for the quotient $\frac{c+di}{a+bi}$ can be obtained by multiplying the numerator and denominator by the number $a - bi$ which is the conjugate of the denominator. For example,

$$\frac{5+i}{1+i} = \frac{(5+i)(1-i)}{(1+i)(1-i)} = \frac{6-4i}{2} = 3 - 2i.$$

7.3. Trigonometric Form of a Complex Number

It is natural to represent the complex number $\alpha = a + bi$ by a point in the plane, assuming a and b to be its coordinates. Then every complex number is associated with a

point and every point in the plane is associated with a complex number. Real numbers are represented by points with zero ordinates, i.e. points which lie on the abscissa axis. The representations of pure imaginary numbers bi lie on the axis of ordinates. The number 0 corresponds to the origin of coordinates.

The plane in which complex numbers are represented is called a *plane of a complex variable.* The abscissa axis is called a *real axis* and the axis of ordinates is called an *imaginary axis* in full accord with the names of the numbers whose representations lie on these axes.

It is also convenient to associate every complex number with a vector emanating from the origin and terminating at the point which represents that number (i.e. the radius vector of that point). The components a and b of the complex number $a + bi$ are evidently the projections (algebraic, with due account of the sign) of the vector on the coordinate axes. As is known, the projection of the sum of vectors (in the sense of vector addition) on any axis is equal to the sum of the projections of the terms. Therefore the sum of the vectors representing the complex numbers α and β is a vector representing the sum $\alpha + \beta$ of these numbers since the components of the number $\alpha + \beta$ are equal to the sums of the respective components of the terms.

The length of the radius vector of the point which represents the complex number α, equal to the distance from the point to the origin, is called a *modulus*, or *absolute value*, of this number and is designated as $|\alpha|$. It is clear that $|\alpha| \geqslant 0$ and $|\alpha| = 0$ only if $\alpha = 0$. The angle which the direction of the radius vector representing the number α makes with the real axis is the *argument* of the number α and is designated as $\arg \alpha$. Note that $\arg \alpha$ has sense only if $\alpha \neq 0$, for the number 0 the argument is meaningless.

The positive direction of reckoning the argument of a complex number is that from the positive real half-axis to the positive imaginary half-axis, i.e. counterclockwise, when the axes are in their ordinary positions.

Just as, in general, an angle between two directions, the argument of a complex number is not uniquely defined but only with an accuracy to within a multiple of 2π. Sometimes it proves to be convenient to impose some restrictions on the

choice of the argument, but more often it is not convenient. Speaking of the argument of a complex number, we shall assume some one of its values, no matter which. Now if we have to choose some definite value, we do it with the aid of an appropriate description (say, "we take the least nonnegative value of the argument").

The modulus $r = |\alpha|$ and the argument $\varphi = \arg \alpha$ of the complex number $\alpha = a + bi$ are connected with its components by the formulas

$$a = r \cos \varphi, \quad b = r \sin \varphi.$$

These formulas follow directly from the definition of the cosine and sine functions of any angle. It is clear that $r = \sqrt{a^2 + b^2}$, $\cos \varphi = a/r$, $\sin \varphi = b/r$. These formulas define the modulus and the argument from the given a and b. We can use the formula $\tan \varphi = b/a$ for $a \neq 0$ to find the argument. But this formula defines φ only with an accuracy to within a multiple of π (i.e. half a revolution) and not to within a multiple of 2π. This necessitates an additional choice between the two values of φ in the opposite quarters. We choose φ whose cosine (or sine) coincides with a in sign (b respectively).

Substituting for the components of the complex number $\alpha = a + bi$ their expressions in terms of the modulus and the argument, we obtain

$$\alpha = r (\cos \varphi + i \sin \varphi).$$

This form of notation of a complex number is known as *trigonometric* notation.

Examples:

$$1 = 1 (\cos 0 + i \sin 0),$$

$$-1 = 1 (\cos \pi + i \sin \pi),$$

$$i = 1 \left(\cos \frac{\pi}{2} + i \sin \frac{\pi}{2}\right),$$

$$1 - i = \sqrt{2} \left(\cos \left(-\frac{\pi}{4}\right) + i \sin \left(-\frac{\pi}{4}\right)\right),$$

$3 + 4i = 5 (\cos \varphi + i \sin \varphi)$, where φ is an angle of the first quarter whose cosine is equal to 3/5.

7.4. Multiplication of Complex Numbers in Trigonometric Form

Let $\alpha_1 = r_1 (\cos \varphi_1 + i \sin \varphi_1)$ and $\alpha_2 = r_2 (\cos \varphi_2 + i \sin \varphi_2)$. Then

$$\begin{aligned}\alpha_1\alpha_2 &= r_1r_2 (\cos \varphi_1 \cos \varphi_2 - \sin \varphi_1 \sin \varphi_2) \\ &\quad + i (\sin \varphi_1 \cos \varphi_2 + \cos \varphi_1 \sin \varphi_2) \\ &= r_1r_2 (\cos (\varphi_1 + \varphi_2) + i \sin (\varphi_1 + \varphi_2)).\end{aligned}$$

Thus, in the trigonometric notation, $\alpha_1\alpha_2$ has a modulus r_1r_2 and argument $\varphi_1 + \varphi_2$, i.e. the modulus of the product of two complex numbers is equal to the product of the moduli of the factors and the argument of the product (one of the values of the argument, to be more exact) is equal to the sum of the arguments of the factors. In the literal notation

$$| \alpha_1\alpha_2 | = | \alpha_1 | \cdot | \alpha_2 |, \quad \arg (\alpha_1\alpha_2) = \arg \alpha_1 + \arg \alpha_2.$$

These rules can be extended to the product of any number of factors, namely,

$$| \alpha_1\alpha_2 \ldots \alpha_k | = | \alpha_1 | \cdot | \alpha_2 | \cdot \ldots \cdot | \alpha_k |,$$

$$\arg (\alpha_1\alpha_2 \ldots \alpha_k) = \arg \alpha_1 + \arg \alpha_2 + \ldots + \arg \alpha_k.$$

Indeed, these formulas hold true for $k = 2$. We assume that they hold true for the product of $k - 1$ factors and obtain

$$\begin{aligned}|\alpha_1\alpha_2 \ldots \alpha_k| &= |\alpha_1| \cdot |\alpha_2 \ldots \alpha_k| = |\alpha_1| \cdot |\alpha_2| \ldots |\alpha_k|, \\ \arg (\alpha_1\alpha_2 \ldots \alpha_k) &= \arg \alpha_1 + \arg (\alpha_2 \ldots \alpha_k) \\ &= \arg \alpha_1 + \arg \alpha_2 + \ldots + \arg \alpha_k.\end{aligned}$$

In both chains of equalities the last transition is made on inductional assumptions.

In the formula

$$\begin{aligned}r_1 (\cos \varphi_1 + i \sin \varphi_1) r_2 (\cos \varphi_2 + i \sin \varphi_2) \ldots r_k (\cos \varphi_k + i \sin \varphi_k) \\ = r_1r_2 \ldots r_k (\cos (\varphi_1 + \varphi_2 + \ldots + \varphi_k) \\ + i \sin (\varphi_1 + \varphi_2 + \ldots + \varphi_k)),\end{aligned}$$

we assume that all the factors are equal so that $r_1 = r_2 = \ldots = r_k = r$, $\varphi_1 = \varphi_2 = \ldots = \varphi_k = \varphi$. We obtain

$$(r (\cos \varphi + i \sin \varphi))^k = r^k (\cos k\varphi + i \sin k\varphi).$$

For $r = 1$ we get the famous *De Moivres's formula*

$$(\cos \varphi + i \sin \varphi)^k = \cos k\varphi + i \sin k\varphi.$$

We have derived this formula on the assumption that k is a positive integer. Let us show that it remains valid for $k = 0$ and for a negative integer k, assuming that $\alpha^0 = 1$ and $\alpha^{-m} = 1/\alpha^m$ for complex numbers as well as for real ones.

For $k = 0$ the formula turns into a true equality:

$$(\cos \varphi + i \sin \varphi)^0 = \cos 0 + i \sin 0 = 1.$$

Now we set $k = -m$ assuming m to be a positive integer, and then

$$\begin{aligned}(\cos \varphi + i \sin \varphi)^k = (\cos \varphi + i \sin \varphi)^{-m} &= \frac{1}{(\cos \varphi + i \sin \varphi)^m} \\ &= \frac{1}{\cos m\varphi + i \sin m\varphi} = \frac{\cos m\varphi - i \sin m\varphi}{\cos^2 m\varphi + \sin^2 m\varphi} \\ &= \cos (-m)\varphi + i \sin (-m)\varphi \\ &= \cos k\varphi + i \sin k\varphi.\end{aligned}$$

Thus De Moivre's formula holds true for all integral values of k.

De Moivre's formula is a convenient means for transforming expressions containing trigonometric functions. We shall consider some examples.

Example 1. Express $\tan 5\varphi$ in terms of $\tan \varphi$.

Solution. We have a relation $\cos 5\varphi + i \sin 5\varphi = (\cos \varphi + i \sin \varphi)^5$. Employing the binomial theorem, we obtain

$$\begin{aligned}\cos 5\varphi + i \sin 5\varphi = &\cos^5 \varphi + 5i \cos^4 \varphi \sin \varphi \\ &- 10 \cos^3 \varphi \sin^2 \varphi \\ &- 10i \cos^2 \varphi \sin^3 \varphi \\ &+ 5 \cos \varphi \sin^4 \varphi + i \sin^5 \varphi\end{aligned}$$

(we use the fact that $i^2 = -1$, $i^3 = -i$, $i^4 = 1$, $i^5 = i$). Equating the components to one another, we get

$$\cos 5\varphi = \cos^5 \varphi - 10 \cos^3 \varphi \sin^2 \varphi + 5 \cos \varphi \sin^4 \varphi,$$

$$\sin 5\varphi = 5 \cos^4 \varphi \sin \varphi - 10 \cos^2 \varphi \sin^3 \varphi + \sin^5 \varphi,$$

whence it follows that

$$\tan 5\varphi = \frac{5\cos^4\varphi\sin\varphi - 10\cos^2\varphi\sin^3\varphi + \sin^5\varphi}{\cos^5\varphi - 10\cos^3\varphi\sin^2\varphi + 5\cos\varphi\sin^4\varphi}$$
$$= \frac{5\tan\varphi - 10\tan^3\varphi + \tan^5\varphi}{1 - 10\tan^2\varphi + 5\tan^4\varphi}.$$

(We have divided the numerator and denominator by $\cos^5\varphi$.)

Clearly, in this way we can express the trigonometric functions of a multiple argument in terms of the trigonometric functions of the initial argument.

Example 2. Express $\sin^5\varphi$ linearly in terms of trigonometric functions of multiple arguments.

Solution. We set $\alpha = \cos\varphi + i\sin\varphi$, and then $\alpha^{-1} = \cos\varphi - i\sin\varphi$, $\alpha^k = \cos k\varphi + i\sin k\varphi$, $\alpha^{-k} = \cos k\varphi - i\sin k\varphi$, whence it follows that

$$\cos\varphi = \frac{\alpha + \alpha^{-1}}{2}, \quad \sin\varphi = \frac{\alpha - \alpha^{-1}}{2i},$$
$$\cos k\varphi = \frac{\alpha^k + \alpha^{-k}}{2}, \quad \sin k\varphi = \frac{\alpha^k - \alpha^{-k}}{2i}.$$

We make use of these formulas:

$$\sin^5\varphi = \left(\frac{\alpha - \alpha^{-1}}{2i}\right)^5 = \frac{\alpha^5 - 5\alpha^3 + 10\alpha - 10\alpha^{-1} + 5\alpha^{-3} - \alpha^{-5}}{32i}$$
$$= \frac{(\alpha^5 - \alpha^{-5}) - 5(\alpha^3 - \alpha^{-3}) + 10(\alpha - \alpha^{-1})}{32i}$$
$$= \frac{2i\sin 5\varphi - 10i\sin 3\varphi + 20i\sin\varphi}{32i}$$
$$= \frac{\sin 5\varphi - 5\sin 3\varphi + 10\sin\varphi}{16}.$$

Similarly, any expression of the form $\cos^k\varphi\sin^m\varphi$ can be linearly represented in terms of trigonometric functions of multiple arguments.

7.5. Extracting a Root of a Complex Number

Let n be a natural number. To extract the nth root of a complex number α is to find a complex number (or numbers) β such that $\beta^n = \alpha$. Every number β such that $\beta^n = \alpha$ is the nth *root of* α and is designated as $\sqrt[n]{\alpha}$.

It is clear that if $\alpha = 0$, then the number 0 is the only value of $\sqrt[n]{\alpha}$ and, therefore, we shall concentrate our atten-

tion on the case $\alpha \neq 0$. We write α in the trigonometric form $\alpha = r(\cos \varphi + i \sin \varphi)$ and seek β also in trigonometric form:

$$\beta = R(\cos \theta + i \sin \theta).$$

We can write the relation $\beta^n = \alpha$ in the form

$$R^n(\cos n\theta + i \sin n\theta) = r(\cos \varphi + i \sin \varphi).$$

Equating the moduli and the arguments to one another (taking many-valuedness into account), we find that the last relation is equivalent to the relations

$$R^n = r,$$

$$n\theta = \varphi + 2k\pi, \quad k = 0, \pm 1, \pm 2, \ldots.$$

The number r is positive (since $\alpha \neq 0$) and the required number R must also be positive. It is known that for any positive number there is a unique positive value of the nth root called an *arithmetic value of the root*, and it is customary to write this value in the form of a power with a fractional exponent. Thus $R = r^{1/n}$. The argument θ can be found by means of a simple division:

$$\theta = \frac{\varphi + 2k\pi}{n}, \quad k = 0, \pm 1, \pm 2 \ldots .$$

Thus nth roots of complex numbers exist and are all defined by the formula

$$\beta_k = r^{1/n}\left(\cos \frac{\varphi + 2k\pi}{n} + i \sin \frac{\varphi + 2k\pi}{n}\right) \tag{1}$$

for any $k = 0, \pm 1, \ldots$ (we put the index k in β to emphasize the many-valuedness of $\sqrt[n]{\alpha}$ and the dependence of its values on the parameter k which can assume all integral values).

Let us prove the following important theorem.

Theorem. *There are exactly n values of the nth root of the complex number* $\alpha = r(\cos \varphi + i \sin \varphi)$. *They are given by the formula*

$$\sqrt[n]{\alpha} = r^{1/n}\left(\cos \frac{\varphi + 2k\pi}{n} + i \sin \frac{\varphi + 2k\pi}{n}\right)$$

on the assumption that k runs over the integral values $0, 1, 2, \ldots, n - 1$.

Proof. We shall first show that the formula gives different roots for $k = 0, 1, 2, \ldots, n-1$. We assume that

$$r^{1/n}\left(\cos\frac{\varphi+2k_1\pi}{n}+i\sin\frac{\varphi+2k_1\pi}{n}\right)$$
$$=r^{1/n}\left(\cos\frac{\varphi+2k_2\pi}{n}+i\sin\frac{\varphi+2k_2\pi}{n}\right),$$

where k_1 and k_2 are integers, $0 \leqslant k_1 < k_2 \leqslant n-1$. The last equality is possible if the arguments appearing on the left-hand and right-hand sides are equal or differ by a multiple of 2π. We have

$$\frac{\varphi+2k_2\pi}{n}=\frac{\varphi+2k_1\pi}{n}+2t\pi$$

for an integral t, whence $\frac{k_2-k_1}{n}=t$. The last equality is impossible since t is an integer and $0<\frac{k_2-k_1}{n}\leqslant\frac{n-1}{n}<1$.

We have only to prove that the values of k which do not belong to the interval $0 \leqslant k \leqslant n-1$ do not yield any new values in the formula for $\sqrt[n]{\alpha}$. Let k be any integer. Then k/n is included between two adjacent integers q and $q+1$ and can be equal to one of them. Thus $q \leqslant k/n < q+1$. Then $nq \leqslant k < nq+n$ and the inequalities $0 \leqslant k' < n$ are satisfied for $k' = k - nq$ and, since k' is an integer, $0 \leqslant k' \leqslant n-1$. Furthermore, $k = nq + k'$ and

$$r^{1/n}\left(\cos\frac{\varphi+2k\pi}{n}+i\sin\frac{\varphi+2k\pi}{n}\right)$$
$$=r^{1/n}\left(\cos\left(\frac{\varphi+2k'\pi}{n}+2q\pi\right)+i\sin\left(\frac{\varphi+2k'\pi}{n}+2q\pi\right)\right.$$
$$=r^{1/n}\left(\cos\frac{\varphi+2k'\pi}{n}+i\sin\frac{\varphi+2k'\pi}{n}\right),$$

so that the value of $\sqrt[n]{\alpha}$ for k coincides with that for $k = k'$ which belongs to the interval $0 \leqslant k' \leqslant n-1$. And this completes the proof of the theorem.

Example 1. Find $\sqrt[3]{2+2i}$ (one of the few well-juggled numerical examples).

Solution. We have $2 + 2i = \sqrt{8}\,(\cos 45° + i \sin 45°)$. According to the formula, we have

$$\sqrt[3]{2+2i} = (\sqrt{8})^{1/3}\left(\cos\frac{45°+k\cdot 360°}{3} + i\sin\frac{45°+k\cdot 360°}{3}\right)$$
$$= \sqrt{2}\,(\cos(15°+k\cdot 120°) + i\sin(15°+k\cdot 120°)).$$

It is sufficient to take the values 0, 1, 2 for k to obtain three values:

$$\beta_0 = \sqrt{2}\,(\cos 15° + i\sin 15°),$$
$$\beta_1 = \sqrt{2}\,(\cos 135° + i\sin 135°),$$
$$\beta_2 = \sqrt{2}\,(\cos 255° + i\sin 255°).$$

Taking into account that $\cos 45° = \sin 45° = 1/\sqrt{2}$, we get $\beta_1 = -1 + i$. To calculate β_0 and β_2, we pay attention to the fact that $15° = 45° - 30°$ so that

$$\cos 15° = \cos 45°\cos 30° + \sin 45°\sin 30° = \frac{1}{\sqrt{2}}\left(\frac{\sqrt{3}}{2} + \frac{1}{2}\right),$$

$$\sin 15° = \sin 45°\cos 30° - \cos 45°\sin 30° = \frac{1}{\sqrt{2}}\left(\frac{\sqrt{3}}{2} - \frac{1}{2}\right),$$

and therefore

$$\beta_0 = \frac{\sqrt{3}+1}{2} + i\,\frac{\sqrt{3}-1}{2}, \quad \beta_2 = -\frac{\sqrt{3}-1}{2} - i\,\frac{\sqrt{3}+1}{2}.$$

It should be pointed out in conclusion that there is no reason to prefer one of the n values of the nth root of the complex number to the others. We do not introduce the concept of the "arithmetic value" of the root extracted of a complex number since it is impossible to do this in any natural way.

7.6. Extracting a Square Root of a Complex Number

We can extract a square root of a complex number without resorting to a trigonometric form. We shall derive the algebraic form to perform this operation.

We assume that $x + yi = \sqrt{a + bi}$ and set $b \neq 0$ since only this case is of interest. Then $a + bi = (x + yi)^2 = x^2 - y^2 + 2xyi$, and this is equivalent to the system of

equations

$$x^2 - y^2 = a, \quad 2xy = b,$$

and we are interested only in the real solutions of this system. We know that the problem has solutions, and so we can assume that x and y are solutions of the problem. Then $(x^2 - y^2)^2 = a^2$, $4x^2y^2 = b^2$. Adding these relations together, we get $(x^2 + y^2)^2 = a^2 + b^2$, whence $x^2 + y^2 = \sqrt{a^2 + b^2}$. We must take here the arithmetic values of the root since $x^2 + y^2 > 0$. Comparing the last relation to the first equation of the system, we obtain

$$2x^2 = \sqrt{a^2 + b^2} + a, \quad 2y^2 = \sqrt{a^2 + b^2} - a.$$

In accordance with the sense of the problem, the right-hand sides of the two relations must be nonnegative, and this is true indeed since

$$\sqrt{a^2 + b^2} > \sqrt{a^2} = |a|.$$

From the last relations we find that

$$x = \sigma_1 \sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}}, \quad y = \sigma_2 \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}}.$$

Here we again take the arithmetic values of the roots and σ_1 and σ_2 assume the values ± 1. It is clear that the numbers x and y we have found satisfy the first equation of the system, $x^2 - y^2 = a$. But they must also satisfy the second equation $2xy = b$. This yields

$$2\sigma_1\sigma_2 \sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}} \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}} = b,$$

or, after obvious transformations,

$$\sigma_1\sigma_2 \sqrt{b^2} = b,$$

whence $\sigma_1\sigma_2 = 1$ if $b > 0$ and $\sigma_1\sigma_2 = -1$ if $b < 0$ so that $\sigma_2 = \sigma_1 \operatorname{sgn} b$, where $\operatorname{sgn} b$ denotes the sign of b, i.e. $+1$ if $b > 0$ and -1 if $b < 0$.

This yields the formula

$$\sqrt{a + bi} = \pm \left(\sqrt{\frac{\sqrt{a^2 + b^2} + a}{2}} + i \operatorname{sgn} b \sqrt{\frac{\sqrt{a^2 + b^2} - a}{2}} \right).$$

Example 1.

$$\sqrt{i} = \pm\left(\sqrt{\frac{\sqrt{1+0}+0}{2}} + i\sqrt{\frac{\sqrt{1+0}-0}{2}}\right) = \pm\frac{1+i}{\sqrt{2}}.$$

Example 2.

$$\sqrt{3-4i} = \pm\left(\sqrt{\frac{\sqrt{25}+3}{2}} - i\sqrt{\frac{\sqrt{25}-3}{2}}\right) = \pm(2-i).$$

7.7. An Exponential and a Logarithmic Function of a Complex Variable

The exponential function a^x of the real variable x (to a positive base) is defined in several stages. First it is defined for natural values of x, as the product of equal factors, then for negative integral and nonzero values of x, with the use of the laws $a^{-n} = 1/a^n$ and $a^0 = 1$. Then fractional exponents are considered for which the value of the exponential function is defined by means of the roots: $a^{p/q} = \sqrt[q]{a^p}$. For irrational values the definition is connected with the fundamental concept of analysis, with the limiting process, from the considerations of continuity. We cannot use these techniques when we try to extend an exponential function to complex values of the exponent and we do not know, for instance, what 2^{1+i} is.

A power with a complex exponent to the natural base $e = 2.71828...$ was first introduced by Euler on the basis of the analysis of a number of constructions of integral calculus. Sometimes algebraic expressions which are very much alike give different results when they are integrated, for example,

$$\int \frac{1}{1-x^2}\,dx = \frac{1}{2}\ln\frac{x+1}{x-1} + C,$$

$$\int \frac{1}{1+x^2}\,dx = \arctan x + C.$$

At the same time the second integral here formally results from the first when x is replaced by xi. Hence we can infer that when an exponential function with a complex exponent is well defined, the inverse trigonometric functions are related to logarithms and thus an exponential function is connected with trigonometric functions.

Euler had courage and fantasy enough to give a reasonable definition of an exponential function to the base e, namely,

$$e^{a+bi} = e^a (\cos b + i \sin b).$$

This formula is a definition of an exponential function of a complex variable and cannot be proved. There is sufficient ground, however, to consider it to be intelligent and reasonable. Here is one of the reasons.

We know that if k is a real number, then the function $y = e^{kx}$ satisfies a very simple differential equation $y' = ky$ with the initial condition $y = 1$ for $x = 0$. It is natural to wish that a similar equation should be satisfied by the function e^{kx} for a complex k, i.e. the function $e^{(a+bi)x} = e^{ax} (\cos bx + i \sin bx)$.

It is easy to make sure (assuming that the derivative of the sum is equal to the sum of the derivatives for the function with complex values as well) that the function $y = e^{ax} (\cos bx + i \sin bx)$ satisfies the equation $y' = (a + bi) y$. Indeed,

$$y' = ae^{ax} (\cos bx + i \sin bx) + e^{ax} (-b \sin bx + bi \cos bx)$$
$$= ay + e^{ax} \cdot bi (\cos bx + i \sin bx) = ay + biy = (a + bi) y.$$

In addition, this function has a value $y = 1$ for $x = 0$.

The arguments we have presented substantiate the expediency of this definition of an exponential function.

It is easy to prove that when the values of an exponential function are multiplied, the exponents are added together. Indeed,

$$e^{a_1+b_1 i} \cdot e^{a_2+b_2 i} = e^{a_1} (\cos b_1 + i \sin b_1) e^{a_2} (\cos b_2 + i \sin b_2)$$
$$= e^{a_1+a_2} (\cos (b_1 + b_2) + i \sin (b_1 + b_2))$$
$$= e^{(a_1+b_1 i)+(a_2+b_2 i)}.$$

It is easy to express trigonometric functions in terms of exponential functions with imaginary exponents. Setting $a = 0$ in the definition, we obtain

$$\cos b + i \sin b = e^{bi}.$$

Replacing b by $-b$, we get

$$\cos b - i \sin b = e^{-bi}.$$

Adding these relations together and subtracting them term by term, we find that

$$\cos b = \frac{e^{bi} + e^{-bi}}{2}, \quad \sin b = \frac{e^{bi} - e^{-bi}}{2i}.$$

These are *Euler's formulas.*

It is also easy to give a definition of a natural logarithm of a complex number.

A complex number given in the trigonometric form $\alpha = r(\cos \varphi + i \sin \varphi)$ can be written in the form $re^{\varphi i}$. This form of the notation of a complex number is known as *exponential.* It retains all good properties of the trigonometric form but is more concise. Furthermore, $\alpha = re^{i\varphi} = e^{\ln r} e^{\varphi i} = e^{\ln r + \varphi i}$. It is therefore natural to assume that $\ln \alpha = \ln r + \varphi i$ so that the logarithm of the modulus of the complex number turns out to be the real part of its logarithm and its argument to be its imaginary part. This explains, to a certain extent, the "logarithmic" property of the argument, i.e. the fact that the argument of a product is equal to the sum of the arguments of the factors.

A logarithmic function thus introduced is defined for all complex numbers except for zero. It is only necessary to remember that a logarithmic function is many-valued by virtue of many-valuedness of the argument. We could establish the many-valuedness, say, by choosing a *branch* of the logarithm for which $-\pi < \varphi \leqslant \pi$, but this is not very convenient. In particular, the property of the logarithm that the logarithm of a product is equal to the sum of the logarithms of the factors is only valid when the many-valuedness is taken into account. Thus, for instance, 0 is one of the values of $\ln 1$, πi is one of the values of $\ln (-1)$ since $-1 = \cos \pi + i \sin \pi = e^{\pi i}$. However, $\ln [(-1) \cdot (-1)] = \pi i + \pi i = 2\pi i$. This is one of the values of the logarithm of 1 (since $1 = \cos 2k\pi + i \sin 2k\pi$) but it is different from 0.

We can finally give a definition of an exponential function to an arbitrary nonzero base.

Let α be a nonzero complex number. Then $\alpha = e^{\ln \alpha}$ for any value of $\ln \alpha$. It is therefore natural to assume by definition that $\alpha^\beta = e^{\beta \ln \alpha}$. This is again a many-valued function of α and β by virtue of the many-valuedness of $\ln \alpha$ which is defined with an accuracy to within the term $2k\pi i$.

Let us see, for instance, what i^i is equal to. Since $\ln i = i\left(\frac{\pi}{2} + 2k\pi\right)$, it follows that $i^i = e^{-\left(\frac{\pi}{2} + 2k\pi\right)}$. The result seems to be somewhat paradoxical since all the values of the "very imaginary" expression i^i are real.

EXERCISES TO CHAPTER 7

1. Sum up the following complex numbers:

1. $(3 - 7i) + (6 + 5i)$. 2. $(-2 + 6i) + (-4 - 8i)$.
3. $(-5 + 4i) + (5 + 10i)$. 4. $(1 - i) + (3 + i)$.
5. $3i + (7 - 2i)$. 6. $i + 8i$.

2. Subtract the following complex numbers from each other:

1. $(2 - 4i) - (-7 - i)$. 2. $(-8 + i) - (9 + i)$.
3. $(10 + 2i) - (10 - i)$. 4. $6 - (5 + 2i)$.

3. Multiply the following complex numbers (n is a natural number):

1. $(3 + 2i)(2 + 3i)$. 2. $(-1 - i)(-2 + 2i)$.
3. $(7 - i)(7 + i)$. 4. $(-4 - i)(4 - i)$.
5. $(3 + i) \cdot i$. 6. $i(2 - 4i)$. 7. i^3. 8. i^4. 9. i^{4n+1}.
10. i^{4n+2}. 11. i^{4n+3}. 12. i^{4n}. 13. $(1 + i)^3$. 14. $(1 + i)^4$.

4. Perform the division of the following complex numbers:

1. $\frac{4+6i}{1-i}$. 2. $\frac{10-i}{1+i}$. 3. $\frac{1-2i}{3+i}$. 4. $\frac{-2-3i}{1-2i}$.

5. In accordance with law 4 (see pp. 301-302), every real number a can be regarded as a complex number $a + 0i$. Show that the results of arithmetic operations involving real numbers coincide with the results of the corresponding arithmetic operations involving the same numbers written as complex numbers.

Answer. 327 600.

6. Prove that if $\bar{z}_1$ and $\bar{z}_2$ are complex numbers conjugate to z_1 and z_2 respectively, and $m = z_1 + z_2$ and $n = \bar{z}_1 + \bar{z}_2$, then m and n are conjugate numbers.

7. Let $z_1 = a + bi$ and $z_2 = c + di$. Show that if $\bar{z}_1 \cdot z_2$ is a real number, then $z_2 = m\bar{z}_1$, where m is a real number, and $\bar{z}_1$ is a complex number conjugate of z_1.

8. Show that the numbers $-1 + 3i$ and $-1 - 3i$ satisfy the equation $x^2 + 2x + 10 = 0$.

9. Perform the addition of the following complex numbers and give a geometric representation of the terms and their sum:

1. $(-7 - 2i) + (-3 + 5i)$. 2. $(6 + 4i) + (-6 + 5i)$.
3. $(2 - 4i) + (-10 + 4i)$. 4. $-8i + 2i$.

10. Write the following complex numbers in trigonometric form:
1. $\sqrt{3}+i$. 2. $-1-\sqrt{3}i$. 3. $-i$. 4. -2.
5. $5+2i$. 6. $4-3i$.

11. Perform the multiplication:

1. $3\left(\cos\frac{\pi}{6}+i\sin\frac{\pi}{6}\right)\cdot 2\left(\cos\frac{\pi}{4}+i\sin\frac{\pi}{4}\right)$.
2. $(\cos 20^\circ + i\sin 20^\circ)\cdot 6(\cos 25^\circ + i\sin 25^\circ)$.
3. $(\cos 60^\circ + i\sin 60^\circ)(\cos 120^\circ + i\sin 120^\circ)$.
4. $(\cos(-120^\circ) + i\sin(-120^\circ))(\cos(-150^\circ) + i\sin(-150^\circ))$.

12. Find the arguments of the following complex numbers first with an accuracy to within a multiple of 2π and then indicate the least positive value of the argument:

1. $\cos\frac{\pi}{4}-i\sin\frac{\pi}{4}$. 2. $\sin\frac{\pi}{3}+i\cos\frac{\pi}{3}$.
3. $-\sin\frac{\pi}{6}-i\cos\frac{\pi}{6}$. 4. $-\cos\frac{5}{6}\pi+i\sin\frac{5}{6}\pi$.

13. Use De Moivre's formula to calculate the following expressions:

1. $\left(\cos\frac{\pi}{4}+i\sin\frac{\pi}{4}\right)^6$. 2. $(\cos 15^\circ+i\sin 15^\circ)^{10}$.
3. $\left(2\left(\cos\frac{\pi}{14}+i\sin\frac{\pi}{14}\right)\right)^7$. 4. $\left(3\left(\cos\frac{2}{3}\pi+i\sin\frac{2}{3}\pi\right)\right)^8$.
5. $(1+i)^{16}$. 6. $(-1+i)^8$.

14. Express $\cos 3\alpha$ and $\sin 3\alpha$ in terms of $\cos\alpha$ and $\sin\alpha$ using De Moivre's formula.

15. Extract the root of the following expressions and give a geometric representation of all the values of the root:

1. $\sqrt[5]{\cos 225^\circ+i\sin 225^\circ}$. 2. $\sqrt[3]{\cos 135^\circ+i\sin 135^\circ}$.
3. $\sqrt[4]{\cos 120^\circ+i\sin 120^\circ}$. 4. $\sqrt{\cos\frac{\pi}{2}+i\sin\frac{\pi}{2}}$.
5. $\sqrt[3]{-1}$. 6. $\sqrt[4]{1}$. 7. $\sqrt[4]{i}$. 8. $\sqrt[3]{-i}$.

16. Factor the following expressions in the domain of complex numbers:

1. x^2+4. 2. x^4-16. 3. x^2+3-4i. 4. x^2+i.

17. Show that in the domain of complex numbers a quadratic equation with real coefficients and with a negative discriminant has two solutions and those solutions are complex conjugate numbers.

18. Solve the following quadratic equations:

1. $x^2-4x+5=0$. 2. $x^2-8x+7=0$.
3. $3x^2+10x+9=0$. 4. $4x^2-14x+13=0$.

19. Write the following complex numbers in exponential form:

1. $e\left(\cos\frac{\pi}{2}+i\sin\frac{\pi}{2}\right)$. 2. $-e\cdot i$. 3. $e(\cos 2\pi+i\sin 2\pi)$.

4. $-e$. 5. $\cos\frac{\pi}{3}+i\sin\frac{\pi}{3}$. 6. $-\frac{\sqrt{3}}{2}+\frac{1}{2}i$.

7. $2(\cos 3+i\sin 3)$. 8. $3+4i$.

20. Find the logarithms of the numbers given in the preceding exercise. Give two or three specific values of the logarithms of these numbers.

Chapter 8

ELEMENTS OF COMBINATORICS AND THE PROBABILITY THEORY

8.1. Simple Combinatorics Problems

Assume that we have several collections of objects and are to compose combinations of these objects satisfying some conditions. Problems involving the calculation of the number of possible combinations are called *combinatorial problems.* For example, in how many ways can we arrange three different objects in a series? The answer is obvious here. We can do this in six ways since it is easy to make possible combinations: 1, 2, 3; 1, 3, 2; 2, 1, 3; 2, 3, 1; 3, 1, 2; 3, 2, 1 (the figures 1, 2 and 3 played the part of "different objects"). But the following well-known combinatorics problem is already rather difficult: how many "lucky" numbers are there among 1 000 000 six-digit numbers of tramway tickets (the number is assumed to be "lucky" if the sum of the first three digits is equal to the sum of the last three digits)? Here the answer 55 252 is not at all obvious and it can be obtained by a direct sorting out with the aid of a powerful computer. Here are solutions of several simple combinatorics problems.

Problem 1. On the number of samples drawn from several sets. We have two sets of objects, the first containing m elements and the second, n elements. How many pairs of elements can we compose selecting one element from each set? For instance, we can compose fifteen pairs

$$\{I, 1\},\ \{I, 2\},\ \{I, 3\},\ \{I, 4\},\ \{I, 5\},\ \{II, 1\},$$
$$\{II, 2\},\ \{II, 3\},\ \{II, 4\},\ \{II, 5\},\ \{III, 1\},$$
$$\{III, 2\},\ \{III, 3\},\ \{III, 4\},\ \{III, 5\}$$

from two sets

$$\{I, II, III\},\ \{1, 2, 3, 4, 5\}.$$

It is easy to solve this problem in a general form. There are m ways of selecting one element from the first set and n ways of selecting one element from the second set. We have got $m \cdot n$ combinations which form pairs.

The problem can be naturally generalized to samples from several sets. Assume that we have k sets containing m_1, m_2, . . ., m_k elements respectively. In how many ways can we form combinations containing one element each from each of the sets? The first two sets yield $m_1 \cdot m_2$ pairs. We can add an element from the third set to every pair of this kind in m_3 ways so that the number of samples of threes is equal to $m_1 \cdot m_2 \cdot m_3$. In m_4 ways we can add an element from the fourth set to every three elements chosen, so that the number of samples of fours is equal to $m_1 \cdot m_2 \cdot m_3 \cdot m_4$. Reasoning in the same manner, we shall find out that the number of samples each containing one element from all the k sets is equal to $m_1 \cdot m_2 \cdot m_3 \cdot \ldots \cdot m_k$.

Problem 2. Permutations. In how many ways can we arrange n elements in a certain order?

We can assume without losing generality that the elements we have to arrange are numbers from a natural series 1, 2, . . ., n. We shall first consider "small" special cases. For $n = 2$ we have two arrangements (permutations): (1, 2) and (2, 1). As we have seen above, for $n = 3$ we can have six arrangements. We shall not sort out all permutations for $n = 4$ but shall reason as follows. The first place can be occupied by one of the four elements. With every choice of the first element, the other three elements in different orders can occupy the remaining three places so that there are six different arrangements for the fixed first element. Consequently, the total number of permutations is $6 \cdot 4 = 24$. Such a reasoning is of a general nature. We denote the number of permutations of $(n - 1)$ elements by P_{n-1} and the number of permutations of n elements by P_n. In every permutation of n elements the first place can be occupied by one of the elements 1, 2, . . ., n so that there are n possibilities of choosing the first element. With every choice of this kind, there are P_{n-1} possibilities of arranging the remaining $n - 1$ elements in the other $n - 1$ places since these arrangements differ only by the order of the elements. We have thus arrived at a relation $P_n = n \cdot P_{n-1}$ from which we successively find, proceeding from the fact that $P_2 = 2$, that $P_3 = 3 \cdot 2 = 3!$, $P_4 = 4 \cdot P_3 = 4 \cdot 3 \cdot 2 = 4!$, $P_5 = 5 \cdot P_4 = 5 \cdot 4 \cdot 3 \cdot 2! = 5!$ and so on. We assume that $P_{n-1} = (n - 1)!$ and obtain $P_n = n \cdot P_{n-1} = n!$.

Reasoning in this way, we practically solve by induction the formula $P_n = n!$, which is a solution of the problem.

Problem 3. Arrangements. In how many ways can we select m elements from a given set containing n elements and arrange them in a certain order? Such arrangements consisting of m elements chosen from the given n elements and differing either by the elements themselves or by their order, or both by the elements and their order are known as *arrangements* and it is customary to denote their number by A_m^n (read "A arrangements of n elements taken m at a time"). For instance, four elements 1, 2, 3, 4 ($n = 4$) taken two at a time ($m = 2$) can be arranged in 12 ways:

$$(1, 2),\ (2, 1),\ (1, 3),\ (3, 1),\ (1, 4),\ (4, 1)$$
$$(2, 3),\ (3, 2),\ (2, 4),\ (4, 2),\ (3, 4),\ (4, 3).$$

Thus $A_2^4 = 12$. How can we calculate A_m^n? We do the following. We take all permutations of n elements and "cut off" the first m elements for each of them discarding the other $n - m$ elements. The arrangements of n elements taken m at a time remain but every arrangement will be encountered as many times as the number of permutations that can be obtained from the discarded $n - m$ elements, i.e. P_{n-m} times. Thus the number of arrangements of n elements taken m at a time is P_{n-m} times smaller than the number of permutations of n elements, i.e.

$$A_m^n = \frac{P_n}{P_{n-m}} = \frac{n!}{(n-m)!} = n \cdot (n-1) \cdot \ldots \cdot (n-m+1).$$

We can present other arguments. We can make all permutations of n elements using the arrangements of these n elements taken m at a time. We shall attach to each of these arrangements the elements which we discarded. Then each arrangement yields P_{n-m} permutations of n elements according to the number of ways in which the remaining $n - m$ elements can be arranged. Thus the number of permutations of n elements is P_{n-m} times larger than the number of arrangements of n elements taken m at a time, i.e.

$$P_n = A_m^n \cdot P_{n-m},$$

whence follows the formula we have derived for the number of arrangements of n elements taken m at a time.

Problem 4. Combinations. In how many ways can we choose a subset of m elements from a set of n elements? Subsets of this kind, which differ only by the elements chosen with no regard to their mutual positions are called *combinations.* Thus, for instance, we can make six combinations, two elements in each, out of four elements 1, 2, 3, 4: {1, 2}, {1, 3}, {1, 4}, {2, 3}, {2, 4}, {3, 4}. (The braces emphasize the difference between combinations and arrangements; the arrangements (1, 2) and (2, 1) are considered to be different although they are composed of the same elements, whereas {1, 2} and {2, 1} constitute the same arrangement.) The number of combinations of n elements taken m at a time is designated as C_m^n or $\binom{n}{m}$, say, $C_2^4 = 6$. Combinations are evidently closely related to arrangements, namely, if, in every arrangement, we disregard the order of the elements, i.e. do not distinguish between the arrangements differing only by the order of the elements, we shall get all combinations. Conversely, if we make various permutations in every combination, we get all arrangements. Therefore, there holds a formula

$$A_m^n = C_m^n \cdot P_m,$$

or, if we replace A_m^n and P_m by their values, $\frac{n!}{(n-m)!} = C_m^n \cdot m!$, whence we get

$$C_m^n = \frac{n!}{(n-m)!\, m!}.$$

We have found that the number of combinations is equal to the binomial coefficient, i.e. to the coefficient of $a^{n-m} \cdot x^m$ in a polynomial equal to $(a + x)^n$. This coincidence is not accidental. The *binomial theorem*, i.e. the formula used to expand $(a + x)^n$ in the powers of x, can be easily set up from considerations of combinatorics, namely,

$$(a + x)^n = (a + x)(a + x) \cdot \ldots \cdot (a + x) \quad (n \text{ factors}).$$

To multiply several polynomials, it is sufficient to multiply their terms, one from each polynomial in all possible combinations, and sum up the results. Here every factor contains two summands. From each factor we can take either the first term a or the second term x. If we take the first term $n - m$ times and the second term m times, then the

product is $a^{n-m} \cdot x^m$. But there are exactly C_m^n ways of choosing the second term m times from n binomial expressions so that the monomial $a^{n-m} \cdot x^m$ enters into $(a + x)^n$ with a coefficient C_m^n.

Here are some examples related to the questions considered.

Example 1. In how many ways can we arrange five black and five white draughts in a series?

Solution. There are ten places in the series we consider. If we know the numbers of places occupied by five black draughts, then we know all the arrangements. Consequently, the required number is

$$C_5^{10} = \frac{10 \cdot 9 \cdot 8 \cdot 7 \cdot 6}{1 \cdot 2 \cdot 3 \cdot 4 \cdot 5} = 252.$$

Example 2. In how many ways can we arrange five black, four white and three red chips in a series?

Solution. As distinct from the preceding example, we cannot make a direct use of the derived formulas here, but to solve the problem, it is easy to employ arguments like those we used to derive the formula for the number of combinations. We label the black chips by the numbers 1, 2, 3, 4, 5, the white chips by the numbers 6, 7, 8, 9 and the red chips by the numbers 10, 11, 12. We know that there are 12! ways of arranging twelve objects in a series. But the arrangements of the chips do not change when we permute the chips labelled by 1-5, the chips labelled by 6-9 and those labelled by 10-12. Therefore the number of different arrangements is equal to $\frac{12!}{5! \cdot 4! \cdot 3!} = 27\,720$.

Example 3. What is the coefficient of $x^5y^4z^3$ in $(x + y + z)^{12}$ written as a polynomial?

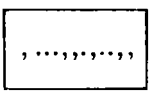
Solution. To calculate the coefficient, we must count the possible ways of choosing the first term five times, the second term four times and the third term three times in the product of twelve factors equal to $(x + y + z)$.

Assuming x to be black, y white and z red, we see that the problem is equivalent to the preceding problem so that the required coefficient is equal to 27 720.

Example 4. Find the number of monomials with coefficient 1 of degree n of m letters.

Solution. Assume that the monomial is equal to $x_1^{k_1} \cdot x_2^{k_2} \cdot \ldots \cdot x_m^{k_m}$. Then $k_1 + k_2 + \ldots + k_m = n$. We put the

following picture into correspondence with this monomial: we take k_1 points, then comma, then k_2 points, then again comma and so on, and, finally, after the $(m-1)$th comma, we take k_m points. There may be zeros among the exponents so that the picture may begin with one or several commas and end with commas, and within the series commas may appear side by side. Thus monomial $x_2^3 \cdot x_4 \cdot x_5^2$ of the letters $x_1, x_2, x_3, x_4, x_5, x_6$ is associated with the picture



It is clear that every picture consisting of n points and $m-1$ commas arranged in any possible way is associated with one monomial of degree n of m letters. Therefore the required number of monomials is equal to the number of pictures of this kind. For every picture we enumerate all the places, occupied by points and commas, by the numbers from 1 to $n+m-1$. Then the picture is completely defined by the indication of the places occupied by commas. There are $m-1$ places of this kind. Consequently, the number of pictures and the number of monomials equal to it is $C_{m-1}^{n+m-1} = C_n^{n+m-1}$. For instance, the number of monomials of degree two of m letters is $C_{m-1}^{2+m-1} = C_2^{m+1} = \frac{(m+1) \cdot m}{2}$, the number of monomials of degree three of three letters is $C_2^{3+3-1} = C_2^5 = 10$.

Example 5. Prove that when dice are thrown three times, the number of ways of getting a score of ten is equal to the coefficient of x^{10} in the polynomial $(x + x^2 + x^3 + x^4 + x^5 + x^6)^3$.

Solution. We must find the number of ways in which the number 10 can be divided into three terms whose order is not indifferent, each term assuming only the values 1, 2, 3, 4, 5, 6. We have the same problem when we multiply the polynomial $x + x^2 + x^3 + x^4 + x^5 + x^6$ by itself three times according to the rule of multiplication of every term of the first factor by every term of the second and the third term. And this solves the problem. Here is the way, for instance, of calculating the required coefficient. The polynomial

$$(x + x^2 + x^3 + x^4 + x^5 + x^6)^3$$

is equal to

$$\frac{x^3(x^6-1)^3}{(x-1)^3} = \frac{x^{21}-3x^{15}+3x^9-x^3}{x^3-3x^2+3x-1}.$$

Performing the division according to the rule of division of polynomials, we obtain

$$x^{18} + 3x^{17} + 6x^{16} + 10x^{15} + 15x^{14} + 21x^{13} + 25x^{12} + 27x^{11} + 27x^{10} + 25x^9 + 21x^8 + 15x^7 + 10x^6 + 6x^5 + 3x^4 + x^3.$$

The coefficient of x^5 yields the number of cases when the score is k in three throws of a die. The required number is 27.

◀ **Example 6.** Prove that the number of "lucky" tramway tickets (with numbers from 000001 to 999999 and 000000) is equal to the coefficient of x^{27} in the polynomial

$$(1+x+x^2+\ldots+x^9)^6 = \frac{(1-x^{10})^6}{(1-x)^6}.$$

Solution. In the same way as in Example 5, it is easy to verify that the coefficient of x^m in the expression $(1 + x + x^2 + \ldots + x^9)^3$ is equal to the number of occurrences of the number m as the sum of the first three digits. We can consider the coefficient of x^{-m} in the expression $(1 + x^{-1} + x^{-2} + \ldots + x^{-9})^3$ to be the number of occurrences of the number $-m$ as the sum, taken with the minus sign, of the second group of three digits. Consequently, the coefficient of x^0 (the constant term) in the expression

$$(1 + x + \ldots + x^9)^3 \cdot (1 + x^{-1} + x^{-2} + \ldots + x^{-9})^3$$

is the required number. But

$$1 + x^{-1} + \ldots + x^{-9} = x^{-9}(1 + x + \ldots + x^9),$$

and so

$$(1 + x + \ldots + x^9)^3 \cdot (1 + x^{-1} + \ldots + x^{-9})^3 = x^{-27}(1 + x + \ldots + x^9)^6$$

and the coefficient of x^0 on the left-hand side is equal to the coefficient of x^{27} in

$$(1+x+\ldots+x^9)^6 = \frac{(x^{10}-1)^6}{(x-1)^6} = \frac{(1-x^{10})^6}{(1-x)^6}.$$

We can find this coefficient if we perform the division of the polynomials and determine the numerator and denominator by means of the binomial formula. But it is better to use the means of analysis, namely, first expand

$$f(x)=\frac{1}{(1-x)^6}=(1-x)^{-6}$$

by Maclaurin's formula. Then it is easy to calculate that

$$f'(x)=6(1-x)^{-7},$$
$$f''(x)=6\cdot 7\cdot(1-x)^{-8},\ \ldots,$$
$$f^{(k)}(x)=6\cdot 7\cdot\ \ldots\ \cdot(6+k-1)(1-x)^{-6-k},$$

and so $f^{(k)}(0)=6\cdot 7\cdot\ \ldots\ \cdot(k+5)$ and

$$\frac{f^{(k)}(0)}{k!}=\frac{6\cdot 7\cdot\ \ldots\ \cdot(k+5)}{k!}=\frac{(k+5)!}{5!k!}=C_5^{k+5}.$$

Consequently, from Maclaurin's formula, interrupted after the term with x^{27}, we obtain a relation

$$(1-x)^{-6}=1+C_5^6x+C_5^7x^2+C_5^8x^3+\ \ldots\ +C_5^{k+5}x^k$$
$$+\ \ldots\ +C_5^{32}x^{27}+x^{28}\cdot F(x),$$

where $F(x)$ is a function bounded in the neighbourhood of $x=0$.

Furthermore,

$$(1-x^{10})^6(1-x)^{-6}=(1-C_1^6x^{10}+C_2^6x^{20}-C_3^6x^{30}+C_4^6x^{40}-C_5^6x^{50}$$
$$+C_6^6x^{60})\cdot(1+C_5^6x+C_5^7x^2+\ \ldots\ +C_5^{32}x^{27})$$
$$+(1-x^{10})^6x^{28}F(x).$$

The left-hand side of the relation includes a polynomial $(1+x+\ \ldots\ +x^9)^6$, the first term on the right-hand side is also a polynomial of x and, consequently, the second term is also a polynomial. Since $(1-x^{10})^6\cdot F(x)$ is a function bounded in the neighbourhood of $x=0$, all the terms of the polynomial $(1-x^{10})^6\cdot x^{28}\cdot F(x)$ contain x of degree not lower than 28. It is easy now to find the required coefficient of x^{27}. It is equal to $C_5^{32}-C_1^6\cdot C_5^{22}+C_2^6\cdot C_5^{12}=55\ 252$. ▶

Example 7. The experiment consists in tossing a coin n times and writing the sequence of obtaining heads and tails. What is the number of possible combinations of occurrences

of heads and tails? In how many combinations will heads appear exactly k times?

Solution. Every toss of a coin has two probabilities, heads or tails. We toss the coin n times. Every possible combination of heads and tails is a sample from n two-element sets and, consequently, the total number of combinations is 2^n (see Problem 1). The second question is similar to the problem on the number of arrangements of n elements taken k at a time, i.e. the number of combinations in which the heads will be obtained exactly k times is equal to C_k^n.

Exercises

1. How many permutations of the elements 1,2, . . ., n are there in which the element 1 does not occupy the first place?
Answer. $P_n - P_{n-1} = (n-1)(n-1)!$

2. How many permutations of the elements 1, 2, . . ., n are there in which the elements 1 and 2 do not occupy their places?
Answer. $P_n - P_{n-1} - P_{n-1} + P_{n-2} = (n^2 - 3n + 3)(n-2)!$

3. How many arrangements of the elements 1, 2, . . ., n taken m at a time ($2 < m < n$) can we make which would not contain the elements 1 and 2 taken together?
Answer. $C_m^n - C_{m-2}^{n-2} = \dfrac{(n-2)!\,(n-1+m)}{m!\,(n-m-1)!}$.

4. How many arrangements of six elements 1, 2, . . ., 6 are there in which 4 occupies the second place and 3 occupies the fourth place?
Answer. $A_3^5 = 120$.

5. How many tramway tickets are there whose numbers are written by four digits?
Answer. 327 600.

6. How many six-digit numbers are there whose notation contains four digits?
Answer. 294 840.

7. Find whether the equation $P_n = 36A_2^{n-1}$ is possible. If it is, then for what n?
Answer. For $n = 6$.

8. Solve the equation $C_5^n = 2C_5^{n-1}$.
Answer. For $n = 10$.

8.2. On Probabilities

In nature and in everyday life we often deal with accidental events, i.e. with situations whose outcome cannot be forecast. Thus, tossing a coin, we cannot guess whether it will come up heads or tails. Similarly, tossing a brick one

of whose faces is coloured, we cannot say whether it will come up the coloured side or not.

When we deal with random events, we come to an idea of probability of an event. We often use these words. Thus we perfectly understand the sense of the following statement: "This event is probable and that event is very improbable". We also use such expressions as "hundred percent probability" and "probability is equal to zero" when we want to emphasize the fact that we are sure of the occurrence of an event or, conversely, of a nonoccurrence of an event. The use of the term "probability" and expressions like those cited above is based on our experience and intuitive notion of probability (possibility) of occurrence of this or that event. Even the fact that we do not know the exact sense of the word "probability" does not prevent us from comparing the probabilities of various events or phenomena.

For instance, if only one face of a brick is coloured, then it is clear that when the brick is tossed, the occurrence of an uncoloured face is more probable. It is also clear that it is less probable that it will come up heads twice in two tosses than the occurrence of a head in one toss.

A hundred occurrence of heads in hundred tosses is a very improbable event but still possible.

The aim of the probability theory is a reasonable quantitative measurement of probabilities of random events and the construction on this basis of a mathematical theory which makes it possible to forecast the processes connected with random events.

Note first of all that the quantitative measurement of the probability of a single random event has no sense. Neither can we answer the question whether the lightning will strike a certain tree during the next thunderstorm or the coin will come up heads 50 times in 50 tosses. Clearly, both these events are next to impossible but there is no reason to consider one of them to be more probable than the other. Thus it is reasonable to speak of the value of probability of a random event only for events which result from experiments admitting a manyfold repetition under the same conditions (say, tossing a coin) or of events occurring at the same time as other events which are under the same conditions (say, radioactive decay of the atoms of radium).

We shall see later on that the so-called *law of large num-*

bers obtains when the probability of an event is determined reasonably. The law states that when an experiment is repeated n times and as a result an event occurs with probability p, then the frequency of the event, i.e. the ratio of the number k of occurrences of the event to the number n of trials, is close, in a certain sense, to the probability, to be more precise, the probability of a significant deviation of the frequency of the event from its probability is small in a large number of trials and decreases with an increase in the number of trials. As we say, the *statistical regularity* of the event is observed in this case.

Let us consider a simple example. We naturally assign probability 1 (a hundred percent probability) to a certain event and probability 0 to an impossible event. In the experiment with tossing a coin, provided that the coin is unbiased, we must consider the probabilities of the coin coming up heads or tails to be equal so that we must assign probability 1/2 to them. Assume that in the experimental verification of the law of large numbers, we tossed a coin 20 times and observed heads 11 times. Then the frequency of its occurrence is equal to $11/20 = 0.55$. We know from literature the results of similar experiments carried out on a considerably larger scale. Thus the well-known French naturalist Buffon (the eighteenth century) observed heads 2048 times in 4040 tosses, i.e. the coin came up heads with the frequency of $\frac{2048}{4040} \cong 0.5069$. The English statistician Pearson described series of tosses performed 12 000 and 24 000 times. In the first case he observed heads 6019 times so that in that experiment the frequency was $\frac{6019}{12\,000} \cong 0.5016$ and in the second case he observed heads 12 012 times, i.e. the frequency was $\frac{12\,012}{24\,000} = 0.5005$. Thus the frequency of occurrences of heads proved to be close to the probability of 0.5, and the larger the series of trials the closer it was.

Let us consider one more example. We have a brick whose faces are coloured as follows: three faces are red, two faces are blue and one face is yellow. We throw the brick onto the table at random. What can we assume to be a natural probability of the brick coming up a red, a blue, a yellow face? There are six possibilities here for the upper face

which we can naturally consider to be equally probable (we threw the brick at random) so that we must consider the probability of the brick coming up each face to be 1/6. There are three red faces and so it can be observed in three cases out of the six and so we must consider the probability of the brick coming up a red face equal to $3/6 = 1/2$. There are two blue faces and so the probability of the brick coming up a blue face is $2/6 = 1/3$, and there is one yellow face and its probability is 1/6.

The next example is more complicated. If we toss a coin six times, then what is the probability of it coming up heads two, three, four times? In this situation 64 outcomes are possible. There are two possibilities in the first trial for each of which there are also two possibilities in the second trial so that the number of possibilities is doubled. When we toss the coin three times, the possibilities are doubled again and so on, so that 64 outcomes are possible in six trials and it is natural to consider these outcomes to be equally probable. The coin comes up heads k times if we observe heads in some k tosses and tails in the rest $6 - k$ tosses. The numbers of these tosses form various combinations of 6 elements taken k at a time and their number is C_k^6. Thus the coin will come up heads two, three or four times in $C_2^6 + C_3^6 + C_4^6 = 15 + 20 + 15 = 50$ cases and the probability that the heads will be observed two, three and four times is equal to $\frac{50}{64} \cong 0.781$. The other possibilities, i.e. the larger deviations of the frequency $k/6$ from 50 percent, have a small probability $\frac{14}{64} \cong 0.219$.

The notion of probability we spoke of is natural from the point of view of intuition. It is easy to generalize our arguments. In all the examples we related the occurrence or non-occurrence of an event A with a collection of equiprobable "elementary events" $a_1, a_2, \ldots, a_n$ so that the event A occurred when some elementary events, favourable for the event A, were realized and did not occur when the other events (unfavourable for the event A) were realized. The probability of the event A is the ratio of the number of elementary events favourable for the event A to their total number.

At first, in the works of the classical scolars on prob-

ability theory such as Fermat (1601-1665), Pascal (1623-1662) and others, the concept of probability was interpreted exactly this way, i.e. on the basis of calculation of the number of favourable equiprobable elementary events. The research carried out by the classics was to a considerable degree aimed at the investigation of games of chance and so at the beginning the probability theory had a small field of application. However, the classical theory is very simple and can serve as a good model for understanding the laws of probability.

We have pointed out above that if an experiment results in an occurrence of a random event to which it is reasonable to assign some probability, then a statistical regularity obtains in manifold repetitions of the trial, i.e. when the number of trials is increased, the frequency of occurrence of the event tends to stabilization and comes close to probability. Thus the statistical regularity of the occurrence of an event is a necessary condition for the possibility of assigning a definite probability to the event. The main field of application of the probability theory is the research, with the use of mathematical means, of processes connected with statistically regular random events. The probability of statistically regular events is assumed to be the frequency of their occurrence in a sufficiently large number of trials (as statisticians say, for a sufficiently large size of a sample). The theory of mathematical statistics itself (based on the probability theory), which is rather advanced, tells us how large a sample must be for the results to be sufficiently trustworthy. This statistical approach to the concept of probability extends the field of application of the theory as compared to the classical approach since it is not always possible to connect statistically regular events with the classical scheme of elementary equiprobable events.

There is no impassable precipice between the classical and statistical approaches to the concept of probability. We can make them closer by means of the following nonstrict reasoning. Assume that we calculate the frequency of occurrence of an event A proceeding from a series of n trials. We agree to consider the trials to be *elementary equiprobable events* and assume those and only those of them for which A occurs to be *favourable for* A. On this assumption the "classical probability" of the event A coincides with the frequency

of its occurrence. This assumption is rather arbitrary and will produce somewhat different results for different series (recall the results of the experiments with tossing a coin). Only for a sufficiently large series will we have plausible results. We repeat that these arguments do not establish an identity between two approaches to the concept of probability but only make them closer.

The necessity of applying the probability theory to more general problems of natural sciences made it necessary to reconsider the fundamentals of the theory. The process of reconsideration was gradual and was more or less completed in the paper (1933) of Academician Kolmogorov, a contemporary Soviet scientist. His substantiation of the principal concepts of the probability theory is constructed on a system of axioms. The classical substantiation is covered by Kolmogorov's theory as a special case. In particular, the theory envelops situations from which there may be infinitely many outcomes. Some problems of this kind were considered much earlier and were solved on the basis of particular ideas (we shall get acquainted with one problem of this kind in 8.13, without going into theory).

8.3. Addition of Probabilities

Assume that A_1 and A_2 are two events related to one and the same set of elementary events. We say that they are *mutually exclusive* (*incompatible*) events if the occurrence of one of them makes the other event impossible. This means that there is no elementary event which would be favourable for both events, A_1 and A_2. The sum of the events A_1 and A_2 (incompatible or compatible) is the event $A_1 + A_2$ consisting in the occurrence of A_1 or A_2.

Theorem. *The probability of the sum of mutually exclusive events A_1 and A_2 is equal to the sum of their probabilities.*

Proof. Assume that $a_1, a_2, \ldots, a_n$ is a system of equiprobable elementary events connected with the events A_1 and A_2. Let k_1 elementary events be favourable for the event A_1 and k_2 elementary events be favourable for the event A_2. Neither of them is favourable for both events A_1 and A_2 at the same time. Consequently, $k_1 + k_2$ elemen-

tary events are favourable for the event $A_1 + A_2$ and, hence,

$$\mathsf{P}(A_1 + A_2) = \frac{k_1 + k_2}{n} = \frac{k_1}{n} + \frac{k_2}{n} = \mathsf{P}(A_1) + \mathsf{P}(A_2).$$

Here $\mathsf{P}(A)$, where A is an event, denotes its probability. We shall use this designation in what follows.

We have proved the theorem on the addition of probabilities proceeding from the classical definition of probability. In the axiomatic substantiation of the probability theory, the theorem on the addition of probabilities is one of the axioms.

The theorem on the addition of probabilities can be easily generalized to the case when the events A_1 and A_2 are compatible. We designate as $A_1 \cap A_2$ an event consisting in the simultaneous occurrence of the events A_1 and A_2. It is known as the intersection of the events A_1 and A_2. The following theorem is true.

Theorem. $\mathsf{P}(A_1 + A_2) = \mathsf{P}(A_1) + \mathsf{P}(A_2) - \mathsf{P}(A_1 \cap A_2)$.

Proof. We designate as k_1, k_2 and l the numbers of elementary events favourable for the events A_1, A_2 and $A_1 \cap A_2$. We see that the number of elementary events favourable for the event $A_1 + A_2$ is equal to $k_1 + (k_2 - l)$ since when taking into account the events favourable for the event A_2, we must exclude the events, taken into account earlier, which are favourable not only for A_2 but also for A_1. The calculation made proves the theorem.

8.4. Multiplication of Probabilities

Random events A and B are *independent* if the probability of the event B does not depend on whether the event A occurred or not. The independence can be seen especially clearly when the events A and B occur in different experiments, independent of one another, related, maybe, to different systems of equiprobable elementary events. A problem concerned with the probability of simultaneous occurrence of independent events A and B, when their probabilities are known, is especially important.

We shall now give a simple example forestalling the investigation of this problem. Assume that we simultaneously throw a die and a coin. Find the probability of occurrence of an even number of points on the die and the heads on the coin. It is natural here to consider pairs of elementary

events:

(1, h), (1, t), (2, h), (2, t), (3, h), (3, t),
(4, h), (4, t), (5, h), (5, t), (6, h), (6, t),

where the digits stand for the number of points on the upper face of the die and the letters h and t stand for heads and tails. We can evidently take the set of pairs for the set of equiprobable elementary events. There are twelve of them and three out of them are favourable for the event being studied: (2, h), (4, h) and (6, h). Thus the required probability is equal to $3/12 = 1/4$.

We shall use now the same arguments to prove the following theorem.

Theorem. *The probability of an event $A \cap B$ consisting in the simultaneous occurrence of independent events A and B is equal to the product of the probabilities of these events.*

Proof. Assume that the event A is connected with equiprobable elementary events $a_1, a_2, \ldots, a_n$, k of which are favourable for the event A so that the probability of the event A is $\mathbf{P}(A) = k/n$ and, respectively, the event B is connected with equiprobable elementary events $b_1, b_2, \ldots, b_m$ among which there are l events favourable for B, so that $\mathbf{P}(B) = l/m$. It is natural to accept the pairs (a_i, b_j) as an elementary event for $A \cap B$, assuming them, by virtue of the independence of the events A and B, to be equiprobable and favourable for $A \cap B$ if and only if a_i is favourable for A and b_j is favourable for B. In accordance with the solution of combinatorics problem 1 on the number of samples chosen from two sets, the number of all pairs (a_i, b_j) is equal to mn, the number of pairs which are favourable for $A \cap B$ being kl. Therefore,

$$\mathbf{P}(A \cap B) = \frac{kl}{mn} = \frac{k}{n} \cdot \frac{l}{m} = \mathbf{P}(A) \cdot \mathbf{P}(B),$$

and this is what we wished to prove.

We have proved the theorem on multiplication of probabilities for independent events within the limits of the classical theory. In modern theory, the relation $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$ is taken as the **definition** *of the independence of the events* A and B.

It is easy to generalize this theorem. We assume that there is a collection of random events $A_1, A_2, \ldots, A_m$ which

are independent in the sense that the probability of each of them does not depend on the occurrence or nonoccurrence of the other ones. Then the probability of the simultaneous occurrence of the events $A_1, A_2, \ldots, A_m$ is equal to the product of probabilities of the events $A_1, A_2, \ldots, A_m$. This fact can be easily verified by induction:

$$\mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}((A_1 \cap A_2) \cap A_3) = \mathbf{P}(A_1 \cap A_2) \cdot \mathbf{P}(A_3)$$
$$= \mathbf{P}(A_1) \cdot \mathbf{P}(A_2) \cdot \mathbf{P}(A_3) \text{ and so on,}$$

$$\mathbf{P}(A_1 \cap A_2 \cap \ldots \cap A_{k-1} \cap A_k)$$
$$= \mathbf{P}((A_1 \cap A_2 \cap \ldots \cap A_{k-1}) \cap A_k)$$
$$= \mathbf{P}(A_1 \cap A_2 \cap \ldots \cap A_{k-1}) \cdot \mathbf{P}(A_k)$$
$$= \mathbf{P}(A_1) \cdot \ldots \cdot \mathbf{P}(A_{k-1}) \cdot \mathbf{P}(A_k).$$

An important case of occurrence of independent random events is observed when an experiment is repeated under the same conditions since the occurrence or nonoccurrence of an event in question in some experiment does not depend on the outcome of the preceding experiments. (It is expedient here to mention a prejudice sometimes stated, namely, it is sometimes assumed that if a coin comes up heads in six trials in succession, then in the next trial the appearance of the heads is less probable than that of the tails. This is an erroneous assumption, of course, since the conditions of trial do not change and do not depend on the outcome of the preceding trials. The feeling that it is less probable to observe heads after their sixfold occurrence is due to the fact that the insufficiently keen intuition in probability problems suggests, instead of the equal, 50-50, probability of occurrence of heads, a comparatively insignificant, 1/128-percent, probability of the sevenfold occurrence of heads, but this probability refers to the beginning of the experiment and not to the moment when six trials have been performed.) Assume that two outcomes A and $\overline{A}$ of an experiment are possible with probabilities p and $q = 1 - p$. Let us calculate the probability of the k-fold occurrence of the event A in n trials. As a result of n trials a series of outcomes may be observed which are various sequences of length n of events A and $\overline{A}$.

If, in a certain series, A occurs k times and, respectively, $\overline{A}$ occurs $n - k$ times, then, by virtue of the theorem on

multiplication of probabilities, the probability of the series is $p^h q^{n-h}$. The numbers of trials resulting in A in such series form combinations of n elements taken k at a time. Since all the different series of outcomes are incompatible, we can use the theorem on the addition of probabilities, and this reduces to the multiplication of the probability of one of the series in question by C_k^n. Thus the sought-for probability is equal to $C_k^n p^h q^{n-h}$. This result is very significant and we shall use it in what follows.

8.5. Application of Probability to Genetics

As a result of experiments, the well-known natural scientist Gregor Mendel (1822-1884) found out a number of regularities when he interbred different sorts of peas. These regularities can be easily explained with the use of the simple probability theory theorems. Their significance is that they form the basis of the theory of heredity as a whole, the so-called *genetics*.

The scheme of application of the probability theory to genetics is based on the following facts. The inheritance of traits depends on special carriers called genes. All the cells of a living organism, except for sex cells, carry the same collection of genes. Genes are parts of chromosomes present in ordinary cells pairwise, and, respectively, genes are also present in cells pairwise at a particular locus in the corresponding chromosomes. In simple cases, each gene of a pair may be in one of two forms (*alleles*). We designate them as A and a. Correspondingly, an organism may have (with respect to a particular gene) three genotypes, AA, Aa and aa. The first and the third genotype are *homozygous* and the second is *heterozygous*.

There are traits defined by one pair of genes. It is the simplest case. There are traits, however, defined by several pairs of genes. We shall consider only the first, simplest, case. Note that more complicated cases can also be investigated by the means of the probability theory.

Sex cells (*gametes*) contain only one gene from each pair. Homozygous individuals produce gametes of only one type and heterozygous individuals of genotype Aa produce gametes of types A and a in equal amounts. A new organism de-

velops from two parental gametes from which it receives genes.

The colour of pea flowers is defined by one gene which has two forms A and a. The peas of genotype AA have red flowers, those of genotype aa have white flowers and the peas of genotype Aa gave pink flowers. Assume that a field is sown with a mixture of red, pink and white peas encountered with frequencies u, $2v$ and w, $u + 2v + w = 1$ (we shall identify frequencies with probabilities because of a large number of peas sown). Then we can tabulate the probabilities of interbreedings:

	AA	Aa	aa
AA	u^2	$2uv$	uw
Aa	$2uv$	$4v^2$	$2vw$
aa	uw	$2vw$	w^2

We could have used the theorem on multiplication of probabilities since the pairs taking part in interbreeding are independent. Every square in the table is, in turn, divided into four squares with the same probabilities in accordance with the possible combinations of gametes. We reduce these possible cases into one table:

AA	AA	AA	Aa	Aa	Aa
AA	AA	AA	Aa	Aa	Aa
AA	AA	AA	Aa	Aa	Aa
Aa	Aa	Aa	aa	aa	aa
Aa	a	Aa	aa	aa	aa
Aa	Aa	Aa	aa	aa	aa

We know the probabilities of all 36 possibilities.

Using now the theorem on the addition of probabilities, we find the probability of combinations AA, Aa and aa.

The probability of combination AA is

$$u^2 + 2uv + v^2 = (u + v)^2.$$

The probability of combination Aa is

$$2uw + 2uv + 2v^2 + 2vw = 2(u + v)(v + w).$$

The probability of combination aa is

$$w^2 + 2vw + v^2 = (w + v)^2.$$

In particular, if we sow half the field with peas with red flowers and half the field with peas with white flowers, i.e. take $u = w = 1/2$, $v = 0$, then approximately 1/4 of the peas will be red, 1/2 will be pink and 1/4 will be white.

It is interesting to point out that if we take the crops with frequencies

$$u_1 = (u + v)^2, \quad 2v_1 = 2(u + v)(w + v), \quad w_1 = (w + v)^2$$

for genotypes AA, Aa and aa, then in the next generation the frequencies will remain unchanged. Indeed,

$$u_2 = (u_1 + v_2)^2 = (u + v)^2 (u + v + w + v)^2 = (u + v)^2 = u_1$$

since $u + 2v + w = 1$. Furthermore,

$$\begin{aligned} v_2 &= (u_1 + v_1)(v_1 + w_1) \\ &= (u + v)(u + v + w + v)(w + v)(u + v + w + v) \\ &= (u + v)(w + v) = v_1, \\ w_2 &= (v_1 + w_1)^2 = (w + v)^2 (w + v + u + v)^2 \\ &= (w + v)^2 = w_1. \end{aligned}$$

As we say, a stable population is obtained already in the first generation.

Sometimes a gene of type A dominates a gene of type a. This means that the organism with genotype Aa does not differ from that with genotype AA (relative to the trait defined by the gene being considered). Thus the green colour of the seeds of peas dominates the yellow colour. If we sow peas of genotypes AA and aa in equal amounts, then three-quarters of the peas of the next generation will have green seeds and only a quarter (genotype aa) will have yellow seeds.

Let us consider the case when a field is sown only with green seeds of peas of genotypes AA and Aa, i.e. in our general scheme we have $w = 0$, $u + 2v = 1$. Then in the next generation the frequency of genotype AA will be $u_1 = (u + v)^2$, the frequency of genotype Aa will be $2v_1 = 2v(u + v)$ and the frequency of genotype aa (peas with yellow seeds) will be $w_1 = v^2$ so that in this generation some peas will have yellow seeds although peas of that genotype were not sown. The portion of peas of genotype AA will increase as compared to the original one:

$$\frac{u_1}{v_1} = \frac{u+v}{v} = 1 + \frac{u}{v}.$$

By excluding peas with yellow seeds when we sow peas next time, we shall increase again the portion of homozygous individuals as compared to heterozygous ones:

$$\frac{u_2}{v_2} = 1 + \frac{u_1}{v_1} = 2 + \frac{u}{v}$$

and so on. Thus, excluding individuals possessing a *recessive* trait from the process of multiplication, we shall get an increase in the homozygous individuals AA as compared to heterozygous ones from generation to generation.

We have considered only the most simple cases of application of the probability theory to problems of genetics. But we can use the same scheme, only with considerable technical complications, to study more intricate cases when a gene has more than two alleles, when a trait is connected with several genes, etc.

8.6. Random Variables

When studying processes connected with random events, we may encounter the so-called random variables. A *random variable* is a quantity to which we cannot assign a definite value but can assign several values with some definite probabilities. It is naturally assumed that a quantity always has a value so that the sum of probabilities assigned to all the values is equal to 1. We shall consider an "ordinary" *deterministic* (having a definite value) variable as a random variable assuming a unique value with probability 1.

For instance, assume that we measure the length of a segment by a ruler with millimetre divisions, but, wishing

to be exact, we try to determine fractions of a millimetre by sight. Assume that we make ten measurements and get 21.3 mm once, 21.4 mm three times, 21.5 mm four times and 21.6 mm twice. (We get different results because of a non-exact placing of the end of the ruler and the inaccuracy of counting by sight.) We can consider the results of measurements to be the values of a random variable with probabilities 0.1, 0.3, 0.4 and 0.2 respectively (we regard the frequencies as probabilities).

But in the probability theory itself it is sometimes expedient to introduce random variables as means of investigation. Thus we can associate the tossing of a coin with a variable assuming the value 1 when the coin comes up heads and the value 0 when it comes up tails. We get a random variable assuming values 1 and 0 with probabilities 1/2.

In a more general case, we can naturally associate a trial, whose outcome is the occurrence or nonoccurrence of an event A with probabilities p and $q = 1 - p$, with a random variable assuming values 1 and 0 with probabilities p and q. When we make such a trial n times, we get a random variable with values $0, 1, \ldots, k, \ldots, n$, and to the value k we assign the probability of occurrence of the event A k times. As we have seen above, this probability is equal to $C_n^k p^k q^{n-k}$. This set of trials is known as *Bernoulli trials*.

Two important characteristics are connected with a random variable, the mean value, or mathematical expectation, of the variable and its variance. If the values of the random variable x are $x_1, x_2, \ldots, x_n$ with probabilities $p_1, p_2, \ldots, p_n$, then the *mean value (expectation)* of x is the number $\mathsf{E}x = p_1x_1 + p_2x_2 + \ldots + p_nx_n$. It is clear that if $p_1 = p_2 = \ldots = p_n = 1/n$, then the mean value of the variable x is equal to the arithmetic mean of its values.

If x is a random variable with values $x_1, \ldots, x_n$ with probabilities $p_1, \ldots, p_n$, then we understand $f(x)$, where f is a function, to be a random variable with values $f(x_1), \ldots, f(x_n)$ with the same probabilities $p_1, \ldots, p_n$. It is clear that

$$\mathsf{E}(ax + b) = a\mathsf{E}x + b$$

and

$$\mathsf{E}(f_1(x) + f_2(x)) = \mathsf{E}f_1(x) + \mathsf{E}f_2(x).$$

Note that $\mathsf{E}(x - c) = 0$ if and only if $c = \mathsf{E}x$ since $\mathsf{E}(x - c) = \mathsf{E}x - c$.

The *variance* (*dispersion*) Var (x) of a random variable x is the expectation $\mathsf{E}\,(x-c)^2$ of the square of deviation $x-c$ of the variable x from its mean value c. We may say that the variance characterizes the spread of values of a random variable about its mean value. The smaller the variance, the closer (in the mean) the values of the variable adjoin its mean value.

The explicit formula for variance is

$$p_1\,(x_1-c)^2+p_2\,(x_2-c)^2+\ldots+p_n\,(x_n-c)^2,$$

where $c=\mathsf{E}x$.

The expectation and variance are characterized by the following extremal property. Let b be a number and $F\,(b)=\mathsf{E}\,(x-b)^2$ be the mean value of the square of deviation of the values of the variable x from the number b. Then $F\,(b)$ has a minimum for $b=\mathsf{E}x$ and that minimum is equal to the variance. Indeed, assume that $c=\mathsf{E}x$. Then $F\,(b)=\mathsf{E}\,(x-c+c-b)^2=(c-b)^2+2\,(c-b)\,\mathsf{E}\,(x-c)+\mathsf{E}\,(x-c)^2=(c-b)^2+\mathrm{Var}\,(x)$ since $\mathsf{E}\,(x-c)=0$.

It is clear from this formula that $F\,(b)$ attains its minimum for $b=c=\mathsf{E}x$ and this minimum is equal to Var (x).

We can indicate another formula for variance, namely,

$$\begin{aligned}\mathrm{Var}\,(x)&=\mathsf{E}\,(x-c)^2=\mathsf{E}\,(x^2-2cx+c^2)\\&=\mathsf{E}\,(x^2)-2c\mathsf{E}\,(x)+c^2=\mathsf{E}\,(x^2)-2c^2+c^2\\&=\mathsf{E}\,(x^2)-c^2=\mathsf{E}\,(x^2)-(\mathsf{E}x)^2.\end{aligned}$$

8.7. The Sum of Independent Random Variables

Assume that we have two random variables: x with values $x_1, x_2, \ldots, x_m$ and with their probabilities $p_1, p_2, \ldots, p_m$, and y with values $y_1, y_2, \ldots, y_n$ and with their probabilities $t_1, t_2, \ldots, t_n$. We consider these random variables to be independent, i.e. assume that the realization of one of them does not affect the probability of values of the other. According to the theorem on multiplication of probabilities

$$p\,((x=x_i)\cap(y=y_j))=p_it_j.$$

The *sum of the random variables* x and y is a random variable with values x_i+y_j with probabilities p_it_j, and if several sums of values are equal, then the probabilities corresponding to them are added together. Thus, if x assumes the values

0, 1, 2 with probabilities 1/4, 1/2, 1/4 and y assumes the values 0, 1, 2, 3 with probabilities 1/6, 1/3, 1/3, 1/6, then their sum $x + y$ assumes the values 0, 1, 2, 3, 4, 5 with probabilities

$$\frac{1}{4}\cdot\frac{1}{6}=\frac{1}{24},\quad \frac{1}{4}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{6}=\frac{1}{6},$$
$$\frac{1}{4}\cdot\frac{1}{3}+\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{4}\cdot\frac{1}{6}=\frac{7}{24},$$
$$\frac{1}{4}\cdot\frac{1}{6}+\frac{1}{2}\cdot\frac{1}{3}+\frac{1}{4}\cdot\frac{1}{3}=\frac{7}{24},$$
$$\frac{1}{2}\cdot\frac{1}{6}+\frac{1}{4}\cdot\frac{1}{3}=\frac{1}{6},\quad \frac{1}{4}\cdot\frac{1}{6}=\frac{1}{24}$$

respectively.

Theorem. *The expectation of the sum of two independent random variables is equal to the sum of the expectations of the terms.*

Proof. Let x assume the values $x_1, x_2, \ldots, x_m$ with probabilities $p_1, p_2, \ldots, p_m$ and y assume the values $y_1, y_2, \ldots, y_n$ with probabilities $t_1, t_2, \ldots, t_n$. Then

$$\begin{aligned}
\mathsf{E}(x+y) &= (x_1+y_1)\,p_1t_1+(x_1+y_2)\,p_1t_2+\ldots+(x_1+y_n)\,p_1t_n\\
&\quad+(x_2+y_1)p_2t_1+(x_2+y_2)\,p_2t_2+\ldots+(x_2+y_n)p_2t_n\\
&\quad\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
&\quad+(x_m+y_1)\,p_mt_1+(x_m+y_2)\,p_mt_2\\
&\quad+\ldots+(x_m+y_n)\,p_mt_n\\
&= x_1p_1t_1+x_1p_1t_2+\ldots+x_1p_1t_n\\
&\quad+y_1p_1t_1+y_2p_1t_2+\ldots+y_np_1t_n\\
&\quad+x_2p_2t_1+x_2p_2t_2+\ldots+x_2p_2t_n\\
&\quad+y_1p_2t_1+y_2p_2t_2+\ldots+y_np_2t_n\\
&\quad\cdots\cdots\cdots\cdots\cdots\cdots\\
&\quad+x_mp_mt_1+x_mp_mt_2+\ldots+x_mp_mt_n\\
&\quad+y_1p_mt_1+y_2p_mt_2+\ldots+y_np_mt_n\\
&= (x_1p_1+x_2p_2+\ldots+x_mp_m)\,(t_1+t_2+\ldots+t_n)\\
&\quad+(p_1+p_2+\ldots+p_n)\,(y_1t_1+y_2t_2+\ldots+y_nt_n)\\
&= \mathsf{E}x+\mathsf{E}y,
\end{aligned}$$

since $t_1+t_2+\ldots+t_n=p_1+p_2+\ldots+p_m=1$.

The proof becomes shorter and more graphic if we use the sign $\sum$ for the sum of the same kind of terms, namely,

$$\begin{aligned}\mathsf{E}(x+y) &= \sum_{\substack{i=1,2,\ldots,m\\ j=1,2,\ldots,n}} (x_i+y_j)\,p_it_j = \sum_{i,j} x_ip_it_j + \sum_{i,j} y_jp_it_j \\ &= \sum_i x_ip_i \sum_j t_j + \sum_i p_i \sum_j y_jt_j \\ &= \mathsf{E}x + \mathsf{E}y, \text{ since } \sum_i p_i = \sum_i t_j = 1.\end{aligned}$$

Theorem. *The variance of the sum of independent random variables is equal to the sum of their variances.*

Proof. We use the designations we have used up till now and introduce new designations $c = \mathsf{E}x$, $d = \mathsf{E}y$. Then

$$\begin{aligned}\mathrm{Var}(x+y) &= \sum_{\substack{i=1,\ldots,m\\ j=1,\ldots,n}} (x_i+y_j-(c+d))^2\,p_it_j = \sum_{i,j} (x_i-c)^2p_it_j \\ &\quad - 2\sum_{i,j}(x_i-c)(y_j-d)\,p_it_j + \sum_{i,j}(y_j-d)^2\,p_it_j \\ &= \sum_i (x_i-c)^2\,p_i \cdot \sum_j t_j \\ &\quad - 2\sum_i (x_i-c)\,p_i \sum_j (y_j-d)\,t_j \\ &\quad + \sum_i p_i \sum_j (y_j-d)^2\,t_j \\ &= \mathrm{Var}(x) + \mathrm{Var}(y), \text{ since } \sum_i p_i = \sum_j t_j = 1,\end{aligned}$$

and each of the sums $\sum_i (x_i-c)\,p_i$ and $\sum_j (y_j-d)\,t_j$ is zero.

The theorems on the expectation and variance of the sum of independent random variables can be naturally generalized to any number of terms.

8.8. Expectation and Variance of a Random Variable Connected with Bernoulli Trials

Let us study the Bernoulli scheme. We consider a series of n independent trials at each of which an event A is observed with probability p and is not observed with probability

$q = 1 - p$. Every separate trial is connected with a random variable assuming the value 1 with probability p and the value 0 with probability q. The sum of n of these random variables assumes the value k when, among the values being added, the number 1 is encountered k times, i.e. when in a series of trials the event A occurs k times. Thus, for the sum of random variables studied, the probability of the value k is equal to the probability of a k-fold occurrence of the event A in n trials. Thus we find that the random variable appearing in the Bernoulli trials is the sum of n independent random variables, each of which assumes the value 1 with probability p and the value 0 with probability $q = 1 - p$. We denote this variable by z. Clearly, $\mathsf{E}z = 1 \cdot p + 0 \cdot q = p$ and $\operatorname{Var}(z) = \mathsf{E}(z^2) - (\mathsf{E}z)^2 = \mathsf{E}(z) - (\mathsf{E}z)^2 = p - p^2 = pq$.

Consequently, for the sum of n independent random variables, each of which is equal to z, i.e. for the Bernoulli trials, the expectation is equal to np and the variance is equal to npq.

◀ Pay attention to the fact that using the theorems on expectation and variance, we have simplified rather complicated sums. Indeed, let x be a random variable related to the Bernoulli scheme. According to the definition, its expectation $\mathsf{E}x$ is equal to $\sum_{k=0}^{n} kC_n^k p^k q^{n-k}$ and its variance is equal to $\mathsf{E}(x^2) - (\mathsf{E}x)^2 = \sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k} - \left(\sum_{k=0}^{n} kC_n^k p^k q^{n-k}\right)^2$. Of course, we can calculate the sums $\sum_{k=0}^{n} kC_n^k p^k q^{n-k}$ and $\sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k}$ directly, by means of algebraic transformations. To calculate $\sum_{k=0}^{n} kC_n^k p^k q^{n-k}$, we can employ the identity $kC_n^k = nC_{n-1}^{k-1}$ and to calculate

$$\sum_{k=0}^{n} k^2 C_n^k p^k q^{n-k} = \sum_{k=0}^{n} k(k-1) C_n^k p^k q^{n-k} + \sum_{k=0}^{n} kC_n^k p^k q^{n-k}$$

we can employ the identity $k(k-1)C_n^k = n(n-1)C_{n-2}^{k-2}$. We recommend the reader to complete the calculations. ▶

8.9. Chebyshev's Inequality

The inequality we shall prove estimates the probability of the deviations of the values of a random variable from its mean value. This inequality was established by the famous Russian mathematician P.L. Chebyshev (1821-1894).

Theorem. *The sum of the probabilities of the values of the random variable x, which deviate from the mean value by not more than $t\sqrt{\operatorname{Var}(x)}$, is greater than $1 - \frac{1}{t^2}$. Here t is an arbitrary number larger than 1.*

In other words, *the probability that the values of a random variable will deviate from its mean value by more than $t\sqrt{\operatorname{Var}(x)}$ is smaller than $1/t^2$.*

Proof. Assume that $x_1, x_2, \ldots, x_n$ are the values of the random variable x and $p_1, p_2, \ldots, p_n$ are their probabilities. Then

$$\operatorname{Var}(x) = p_1(x_1 - c)^2 + p_2(x_2 - c)^2 + \ldots + p_n(x_n - c)^2,$$

where $c = \mathrm{E}x$. We leave in this sum only those terms which correspond to the values $x_1, x_2, \ldots, x_n$ removed from c by more than $t\sqrt{\operatorname{Var}(x)}$. We can assume without losing generality that $x_1, x_2, \ldots, x_m$ are these values. Then

$$\operatorname{Var}(x) \geqslant p_1(x_1 - c)^2 + p_2(x_2 - c)^2 + \ldots + p_m(x_m - c)^2.$$

But $|x_i - c| > t\sqrt{\operatorname{Var}(x)}$ for $i = 1, 2, \ldots, m$ and consequently

$$\operatorname{Var}(x) > p_1 t^2 \operatorname{Var}(x) + p_2 t^2 \operatorname{Var}(x) + \ldots + p_m t^2 \operatorname{Var}(x)$$
$$= t^2 \operatorname{Var} x\,(p_1 + p_2 + \ldots + p_m),$$

whence it follows that $p_1 + p_2 + \ldots + p_m < 1/t^2$.

But $p_1 + p_2 + \ldots + p_m$ is exactly the probability that the values of the random variable will deviate from its mean value more than by $t\sqrt{\operatorname{Var}(x)}$. We have proved the theorem.

8.10. The Law of Large Numbers for the Bernoulli Trials

Let us consider a random variable x related to the Bernoulli trials. Its values are $0, 1, 2, \ldots, k, \ldots n$ with probabilities $C_0^n q^n$, $C_1^n p q^{n-1}$, $\ldots$, $C_k^n p^k q^{n-k}$, $\ldots$, $C_n^n p^n$. Its expected value is np and its variance is npq.

According to Chebyshev's inequality we have

$$P(|k-np| \leqslant t\sqrt{npq}) > 1-\frac{1}{t^2},$$

or, which is the same thing,

$$P\left(\left|\frac{k}{n}-p\right| \leqslant t\sqrt{\frac{pq}{n}}\right) > 1-\frac{1}{t^2}.$$

We set $t\sqrt{pq/n} = \varepsilon$. Then $t = \frac{\varepsilon\sqrt{n}}{\sqrt{pq}}$ and we write the inequality in the form

$$P\left(\left|\frac{k}{n}-p\right| \leqslant \varepsilon\right) > 1-\frac{pq}{\varepsilon^2 n}.$$

For a fixed ε the right-hand side tends to 1 as $n \to \infty$. Thus, when the number n of trials increases, the probability that the frequency k/n of occurrence of the event will deviate from its probability by not more than the preassigned small number ε tends to 1 as $n \to \infty$.

This is the essence of the law of large numbers for the Bernoulli scheme which consists in the fact that the frequency of occurrence of an event in a large series of trials is, in general, close to the probability. The deviations are possible but not very probable.

8.11. Random Walk in the Plane

Assume that in equal time intervals a particle which lies on a straight line jumps to the right or to the left by one scale unit with equal probabilities $p = q = 1/2$. The motion of the particle is called a *random walk in the plane.*

Let us see where the particle may be after n jumps. It may be displaced by n units to the right or to the left, but it is very improbable.

To investigate the problem, we shall compare it to the Bernoulli trials for probability $p = q = 1/2$. We put every occurrence of the event A in the Bernoulli trials into correspondence with a jump of the particle to the right and every nonoccurrence of the event with a jump to the left. Then, after n jumps, the abscissa x of the particle is equal to $k - (n - k) = 2k - n$, so that the random variable x

is connected with the random variable y equal to the number of occurrences of the event A in n trials by the relation $x = 2y - n$. Consequently,

$$\mathsf{E}x = 2\mathsf{E}y - n = 2 \cdot \frac{1}{2} n - n = 0.$$

The result is quite natural, i.e. the position of the particle at the midpoint coincides with its position at the beginning of the walk. The calculation of the variance shows that

$$\operatorname{Var}(x) = \mathsf{E}(x^2) = \mathsf{E}(4y^2 - 4ny + n^2) = 4\mathsf{E}\left(\left(y - \frac{n}{2}\right)^2\right)$$

$$= 4 \operatorname{Var}(y) = 4n \cdot \frac{1}{2} \cdot \frac{1}{2} = n.$$

Chebyshev's inequality yields

$$P(|z| < t\sqrt{n}) > 1 - \frac{1}{t^2},$$

so that the deviation of the walking particle from the middle position by a value appreciably larger than $\sqrt{n}$ is not very probable.

This problem can be regarded as a very rough approximation of a description of a physical phenomenon known as *Brownian motion*. It consists in the fact that small particles, suspended in a liquid, perform random movements, being pushed by liquid molecules which perform a chaotic thermal motion. The motion in a plane (and not in space) may be interpreted as a motion in a thin capillary tube.

In this problem we can regard the abscissa of the randomly walking particle as a random variable resulting from the addition of independent random variables, each of which assumes the values $+1$ and -1 with equal probabilities. We shall get a more accurate description of the walk of a particle if we regard the abscissa of every jump as a more general random variable with the mean value equal to zero. Then we can consider the random variable, yielding the abscissa of the particle after n jumps, to be the sum of n equal independent random variables, each of which defines the value of the jump. However, such a generalization changes but little the quantitative aspect of the phenomenon. If δ is the variance of the value of a jump, then, according to the theorem on the addition of the variances of independent

random variables, the variance of the abscissa after n jumps is equal to $n\delta$ and Chebyshev's inequality yields $P(|z| < t\sqrt{\delta n}) > 1 - \frac{1}{t^2}$, so that the deviation of the particle by a value appreciably larger than $\sqrt{n}$ is not very probable.

The generalized problem we have considered has other interesting applications. Here is one of them. Assume that we perform the addition of n numbers whose approximate values we know with an accuracy to within δ. What can we say about the error of the sum? The ordinary theory of errors states that the error of the sum does not exceed $n\delta$. It is clear, however, that this boundary value is attained only in a special case when all the errors of the terms have maximum values and like signs. From the point of view of the probability theory the estimate $n\delta$ is too high. The matter is that the errors of the terms are identical independent random variables with a zero mean value if the construction of approximations was reasonable (say, by means of rounding-off in the necessary direction). We shall see in the next section that the variance should be considered equal to $\frac{1}{3}\delta^2$. *The error of the sum is a random variable which is the sum of n similar random variables.* Its mean value is 0 since the mean values of the terms are zero, and the variance is equal to $\frac{1}{3}\delta^2 n$. According to Chebyshev's inequality

$$P\left(|z| < \frac{t}{\sqrt{3}}\delta\sqrt{n}\right) > 1 - \frac{1}{t^2},$$

so that the error of the sum exceeding $5\delta\sqrt{n}$, for instance, is smaller than 1/75. Thus the estimate $5\delta\sqrt{n}$ for the error, which is appreciably more accurate for large n than the estimate δn with respect to maximum, is almost valid.

8.12. Random Variables Whose Values are Concentrated in an Interval or Throughout the Real Axis

Random variables indicated in the title are often encountered in the probability theory and its applications. A zero probability is assigned, in general, to separate values of

these random variables and nonzero probabilities are attributed to events consisting in the belonging of the values of a random variable to intervals of the real axis. In such a situation a convenient means of giving a probability characteristic to a random variable is a specification of its distribution function. The distribution function of the random variable ξ is a function $F(x)$ equal to the probability that the values of ξ do not exceed x. A distribution function must exist for any random variable. Assume now that $x_1 < x_2$. Then the event $\xi \leqslant x_2$ is the sum of two incompatible events

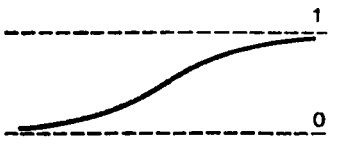


Fig. 8.1

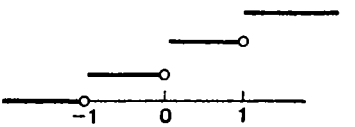


Fig. 8.2

$\xi \leqslant x_1$ and $x_1 < \xi \leqslant x_2$. Therefore the probability $F(x_2)$ of the event $\xi \leqslant x_2$ is the sum of the probability $F(x_1)$ of the event $\xi \leqslant x_1$ and the probability of the event $x_1 < \xi \leqslant x_2$. Thus this latter probability is equal to $F(x_2) - F(x_1)$. Consequently, $F(x_2) - F(x_1) \geqslant 0$ so that $F(x)$ is a nondecreasing function. Furthermore, $\lim\limits_{x \to -\infty} F(x)$ is the limit of the probability that the values of the variable ξ become smaller than any negative number arbitrarily large in absolute value. It is clear that this limit is equal to zero. As $x \to +\infty$, $\lim F(x)$ is the probability that ξ assumes any values unbounded from above, i.e. this limit is equal to 1.

Thus the distribution function of a random variable is a nondecreasing function defined throughout the real axis, which "begins" with zero ($\lim\limits_{x \to -\infty} F(x) = 0$) and "ends" with unity ($\lim\limits_{x \to -\infty} F(x) = 1$). Its graph has, in general, the form shown in Fig. 8.1. If all the values of a random variable are on a finite interval $[a, b]$, then

$$F(x) = 0 \quad \text{for} \quad x < a$$

and

$$F(x) = 1 \quad \text{for} \quad x > b.$$

For a random variable which assumes a finite number of values the concept of a distribution function evidently also has sense. Thus for a random variable equal to -1 with probability 1/4, to 0 with probability 1/2 and to 1 with probability 1/4, the distribution function $\mathsf{F}(x)$ is equal to 0 for $x < -1$, to 1/4 for $-1 \leqslant x < 0$, to 3/4 for $0 \leqslant x < 1$ and to 1 for $x \geqslant 1$. Its graph has the form shown in Fig. 8.2. It is discontinuous and its points of discontinuity are the values of x which are the values of the random variable in question which have nonzero probabilities.

In general, the points of discontinuity of the distribution function of a random variable are numbers which are the values of the random variable with the preassigned nonzero probabilities.

Especially significant are random variables whose distribution function is not only continuous but also differentiable. The differentiability means that for any x there is a limit $\mathsf{F}'(x)$ of the ratio $\frac{\mathsf{F}(x+\Delta x)-\mathsf{F}(x)}{\Delta x}$ as $\Delta x \to 0$. The difference $\mathsf{F}(x+\Delta x)-\mathsf{F}(x)$ is the probability that the random variable assumes values in the interval $(x, x+\Delta x]$ and the existence of $\mathsf{F}'(x)$ means that this probability is almost proportional to the length of the interval, i.e. the distribution function $\mathsf{F}(x)$ is almost linear in the neighbourhood of the point x. The function $\varphi(x) = \mathsf{F}'(x)$ is called the *density of probability*. The probability density $\varphi(x)$ is a nonnegative function for which there is an integral

$$\int_{-\infty}^{\infty} \varphi(x)\,dx = \lim_{\substack{M\to\infty \\ N\to\infty}} \int_{-N}^{M} \varphi(x)\,dx$$

and this integral is equal to 1. A distribution function can be found from the probability density with the use of the formula

$$\mathsf{F}(x) = \int_{-\infty}^{x} \varphi(t)\,dt.$$

The probability that the values of the random variable ξ are in the interval $(a, b]$ (or, what is the same here, in the interval $[a, b]$, since in this situation $\mathsf{F}(x)$ is continuous and the probability that the value of ξ is equal to a is zero) is

equal to

$$\mathsf{F}(b) - \mathsf{F}(a) = \int_a^b \varphi(x)\, dx.$$

For a random variable which has a probability density the expectation and variance are defined in a natural fashion. These definitions are similar to the respective definitions for random variables with a finite number of values, but the sums are replaced by integrals: the expectation

$$\mathsf{E}\xi = \int_{-\infty}^{\infty} x\varphi(x)\, dx$$

and the variance

$$\mathrm{Var}(\xi) = \int_{-\infty}^{\infty} (x-c)^2 \varphi(x)\, dx,$$

where $c = \mathsf{E}\xi$. In this situation an expectation and, all the more so, variance may not exist. But in cases most significant for applications they exist.

Let us consider two important examples. A random variable is *uniformly distributed on the interval* $[a, b]$ if its prob-

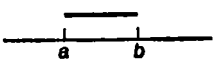


Fig. 8.3

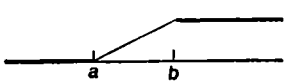


Fig. 8.4

ability density is constant on this interval and is zero outside the interval. The density on the interval must be equal to $\frac{1}{b-a}$ since the density integral along the whole axis must be equal to 1. Figures 8.3 and 8.4 show densities and distribution functions. The expectation of a variable uniformly distributed on $[a, b]$ is

$$\frac{1}{b-a}\int_a^b x\, dx = \frac{1}{b-a}\left(\frac{b^2}{2} - \frac{a^2}{2}\right) = \frac{a+b}{2},$$

and so, as could be expected, it coincides with the midpoint of the interval where the probability density is concentrated.

The variance is equal to

$$\frac{1}{b-a}\int_a^b \left(x-\frac{a+b}{2}\right)^2 dx = \frac{1}{b-a}\left(\frac{1}{3}\left(b-\frac{a+b}{2}\right)^3 - \frac{1}{3}\left(a-\frac{a+b}{2}\right)^3\right) = \frac{1}{12}(b-a)^2.$$

If the density is concentrated on the interval $[-\delta, +\delta]$ (and this is natural to assume for the error of a number given approximately with an accuracy to within δ), then the expectation is 0 and the variance is equal to $\frac{1}{3}\delta^2$.

The so-called *normal law of distribution* is very important for applications. Its density is given by the formula

$$\varphi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-a)^2/(2\sigma^2)}.$$

The corresponding distribution function is expressed in the form of an integral that "cannot be taken", i.e. cannot be expressed in terms of elementary functions. In particular, we cannot present an elementary proof of the fact that $\int_{-\infty}^{\infty} \varphi(x)\,dx = 1$.

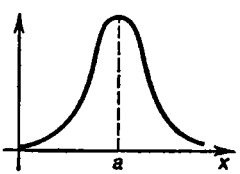


Fig. 8.5

There are expectation and variance for the normal law, namely, the expectation is a and the variance is σ^2. Figure 8.5 shows the graph of density. The bulk of the "mass of probabilities" is in the neighbourhood of the point a. Away from a the density decreases fast and the infinite branches of the graph come very close to the abscissa axis.

The calculation of the values of distribution function $\int_{-\infty}^{x} \varphi(t)\,dt$ by means of the substitution $\frac{t-a}{\sigma\sqrt{2}} = y$ reduces to the calculation of the values of the integral $\int_{-\infty}^{y_0} e^{-y^2}\,dy$. This integral has been tabulated with high accuracy.

The importance of the normal law is that, first, it is true for many random variables encountered in the process of the study of natural phenomena, and, second, it is of especial significance in the investigation of repeated trials made under the same conditions, namely, if $\xi_1, \xi_2, \ldots, \xi_n$ are independent similarly distributed random variables with expectation a and variance σ^2, then the expectation of the arithmetic mean $\overline{\xi} = (\xi_1 + \xi_2 + \ldots + \xi_n)/n$ is a and the variance is σ^2/n. By definition, the variance is the mean value of the squares of deviations of the values of a random variable from the mean so that the square root of the variance is, in a certain sense, the mean (the *root-mean-square*) deviation of the values from the mean. It is therefore natural to compare deviation to the root of the variance. As applied to the variable $\overline{\xi}$, this leads to the consideration of the random variable $\frac{\sqrt{n}}{\sigma}(\overline{\xi} - a)$. The following remarkable theorem proves to hold true.

Limit theorem of probability theory. *The distribution function of the variable* $\frac{\sqrt{n}}{\sigma}(\overline{\xi} - a)$ *converges, as* $n \to \infty$, *to the distribution function of the normal law with density* $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$.

We shall not prove this theorem. Its significance is that when we add independent similarly distributed random variables or, what is almost the same, when we pass to the arithmetic mean, the peculiarities of the initial variable are smoothed off and the distribution becomes more and more like normal. The limit theorem has broad generalizations.

8.13. Buffon's Needle Problem

To conclude this chapter, we shall discuss a problem connected with an experiment which has an infinite number of various outcomes. This is a problem of historic interest and a somewhat unexpected answer to it demonstrates that different divisions of mathematics are interwoven very closely.

Here is the problem. Equispaced parallel straight lines with the distance $2a$ between the neighbouring lines are

drawn in the plane. A needle $2a$ long is thrown at random on the plane. What is the probability that the needle will cut at least one of the lines?

The position of the needle with respect to the system of straight lines is defined by the values of two random variables, first, by the distance y from the midpoint of the needle to the nearest line which may vary from 0 to a, and, second, by the angle φ the direction of the lines makes with that of the needle reckoned from the midpoint k of the nearest

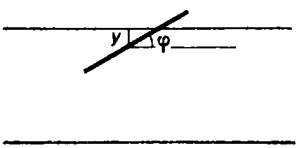


Fig. 8.6

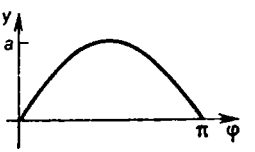


Fig. 8.7

straight line. This angle may assume values from 0 to π. The values satisfying the condition $y \leqslant a \sin \varphi$ are "favourable" for the event we are interested in (Fig. 8.6).

All possible positions of the needle relative to the straight lines are associated with points in the plane with coordinates (φ, y). They fill in a rectangle of length π along the φ-axis and of altitude a along the y-axis. The points corresponding to the "favourable" positions lie between the 0φ-axis of abscissas and the arc of the sine curve $y = a \sin \varphi$ (Fig. 8.7).

Since we have thrown the needle at random, the random variables y and φ must be regarded as independent and uniformly distributed variables on the intervals $[0, a]$ and $[0, \pi]$ of their possible values. Since the distribution is uniform, the probability that y is in the interval $[c_1, c_2]$, $0 \leqslant c_1 < c_2 \leqslant a$, is equal to $\frac{1}{a}(c_2 - c_1)$. Correspondingly, the probability that φ is in the interval $[\alpha_1, \alpha_2]$, $0 \leqslant \alpha_1 < \alpha_2 \leqslant \pi$, is equal to $\frac{1}{\pi}(\alpha_2 - \alpha_1)$. Since y and φ are independent, the probability that the point (φ, y), which defines the position of the needle, lies in the rectangle $c_1 \leqslant y \leqslant c_2$, $\alpha_1 \leqslant \varphi \leqslant \alpha_2$, is equal to the product $\frac{(c_2 - c_1)(\alpha_2 - \alpha_1)}{a\pi}$

of the probabilities that φ and y are in their intervals, i.e. is proportional, with coefficient $\frac{1}{a\pi}$, to the area of the rectangle. Since the plane figure bounded by smooth curves contains itself a figure composed of rectangles and is contained within another similar figure, the areas of the figures differing arbitrarily little from the area of the figure in question, the probability that the point (φ, y) is in the plane figure is proportional to the area of this figure with the same proportionality factor $\frac{1}{a\pi}$. Thus the probability that the needle will cut at least one of the straight lines is equal to the area of the figure bounded by the 0φ-axis and the arc of the sine curve $y = a \sin \varphi$ multiplied by $\frac{1}{2\pi}$. This area is equal to $\int_0^{\pi} a \sin \varphi \, d\varphi = -a \cos \pi + a \cos 0 = 2a$ and, consequently, the required probability is equal to $2/\pi$.

Buffon, who investigated this problem, made a large series of experiments with a needle and calculated the frequency with which the needle cut one of the lines and found an approximate value of the number π assuming that the frequency was close to the probability. The result he obtained was close to the true value of the number π. Thus the number π whose origin is connected with geometry, was approximately calculated by means of a statistical experiment which had nothing to do with the measurement of the length of a circle.

HINTS AND ANSWERS

Chapter 1

4. x is an infinitesimal. **5.** y is not an infinitesimal. **11.** Yes, it will. **12.** No, it will not. **13.** y tends to zero appreciably faster than x. **14.** No, we cannot. **19.** A finite number of them. **21.** No, it will not. **29.** (1) 2. (2) 0. (3) 1.25. (4) 1. **37.** (3) 6. (5) -0.375. **40.** (1) $3x^2$. (3) $\frac{-x^2+1}{(x^2+1)^2}$. (5) $1.5\sqrt{x}$. **43.** (1) $y=2x-1$. (3) $y=0.5$. (5) $y=-4x+4$. (7) $y=x+0.25$. **45.** $y=2x-1$ and $y=6x-9$. **49.** The distance varies according to the law $5t+10$, the velocity is equal to 5. **50.** $l=\frac{9}{8}t$. **51.** $2x\,dx$. **52.** $\frac{dx}{2\sqrt{x}}$. **53.** $6x^5\,dx$. **54.** $\frac{x\,dx}{\sqrt{x^2+1}}$.

Chapter 2

1. (1) $y'=14x^6-15x^4+12x^2-1$. (3) $u'=-16t^3+4t$. (5) $f'=-\frac{3}{z^4}+\frac{1}{z^2}$. (7) $h'=4t^3+\frac{4}{t^5}$. (9) $y'=\frac{-3}{(x-3)^2}$. (11) $y'=\frac{4t}{(t^2+1)^2}$. (13) $f'=\frac{4x^2+10x-11}{(x^2+x+4)^2}$. (15) $g'=15(3u+7)^4$. (17) $S'=4\pi r+4\pi l$. (19) $F'=-14\pi l+6\pi H-16\pi$. (21) $P'=-\frac{2a}{(\alpha-4)^3}+\frac{3b}{(\alpha+3)^4}$. **3.** $y'(x)=x^2(-50x^2+28x-3)$. **5.** $x=3$ and $x=-1$. **7.** (1) $y=18x-21$. **10.** (1) $\frac{3}{3x-7}$. (3) $\frac{2x+2}{x^2+2x}$. (5) $\frac{1}{x\ln 2}$. (7) $\frac{3\log^2 x}{x\ln 10}$. (9) $2x\ln x+x$. (11) $\frac{(x+2)(x+3)+(x+1)(x+2)+(x+1)(x+3)}{(x+1)(x+2)(x+3)}$. (13) $-\frac{x^2+14x-7}{(x^2-x)(x+7)\ln 2}$. (15) $\frac{1-2\ln x}{x^3}$. (17) $\frac{1}{x}$. **12.** (1) $y=x-1$. (3) $y=\frac{1}{2\ln 10}x-\frac{1}{\ln 10}+\log 2$. **15.** (1) $y'<0$ for $x<1.5$. (2) $y'<0$ for $x>2.5$. **16.** (1) $3^x\ln 3$. (3) $(2x-1)e^{x^2-x}$. (5) $2te^t+t^2e^t$. (7) $\frac{1-2z^2}{e^{z^2}}$. (9) $2^{2x-1}\ln 2$. (11) $3\cdot 2^{3t}\ln 2$. (13) $(5x+1)^x\times\left(\ln(5x+1)+\frac{5x}{5x+1}\right)$. **17.** (1) $y=x+1$. (3) $y=x\ln\sqrt{8}+\frac{1}{2}-$

$\ln \sqrt{8}$. **18.** $x = \ln 2$. **19.** (1) $-1.3x^{-2.3}$. (3) $\dfrac{3.5}{\sqrt{7x-1}}$. (5) $\dfrac{-3x+2}{2\sqrt{(4-3x)^3}}$. (7) $\dfrac{-5}{3\sqrt[3]{(-5t+2)^2}}$. (9) $\dfrac{4u-2}{3\sqrt[3]{u^2-u}}$. **20.** (1) $y = \dfrac{1}{4}x + 1$. (3) $y = 3x - 4$. **22.** $x = \dfrac{1}{2\sqrt[3]{2}}$. **23.** (1) $y' = \dfrac{4}{3}$. (3) $y' = -\dfrac{y}{x}$. (5) $y' = \dfrac{y+1}{2-x}$. **24.** (1) 0.

Chapter 3

4. (1) Increasing. (3) Increasing. (5) Decreasing. **5.** (1) Increases on $(-\infty, 1/3]$ and on $[1, \infty)$, decreases on $[-1/3, 1]$. (3) Increases on $[-2, 2]$, decreases on $(-\infty, -2]$ and on $[2, \infty)$. (5) Increases on $[1/e, \infty)$, decreases on $(0, 1/e]$. (7) Increases on $[3, (6+\sqrt{24})/3]$, decreases on $[(6+\sqrt{24})/3, \infty)$. **6.** (7) Increases on $[-1, 0]$ and on $[1, \infty)$, decreases on $(-\infty, -1]$ and on $[0, 1]$, $x = 0$ is a point of maximum, $x = \pm 1$ are points of minimum. (11) Increases on $[1/e, \infty)$, decreases on $(0, 1/e]$, $x = 1/e$ is a point of minimum, $y \to 1$ as $x \to 0$. **8.** (3) The domain of definition of the function is $(-\infty, -1/3]$ and $[1/3, \infty)$. The graph consists of two branches which are symmetric about OY. With an increase in x the branches come close to the straight lines $y = 3x$ and $y = -3x$, respectively, from below and almost merge with them. (5) $x = -2$ is a point of maximum, $x = 1$ is a point of minimum. (7) $x = -1 - \sqrt{2}$ is a point of maximum, $x = -1 + \sqrt{2}$ is a point of minimum. As $x \to -\infty$, the graph comes close to $y = 1$ from above and as $x \to \infty$, the graph comes close to $y = 1$ from below. (9) The domain of definition of the function is $(-1, 2)$. The function increases from $-\infty$ to ∞, $y \to -\infty$ as $x \to -1$ and $y \to \infty$ as $x \to 2$, $y = 0$ for $x = 1/2$. (11) The graph is symmetric about OY, $x = 0$ is a point of minimum. When x's are large in absolute value, the graph differs but little from the graph of the function $y = \sqrt[3]{x^2}$. (13) $x = 1/e^2$ is a point of minimum, $y \to 1$ as $x \to 0$. **10.** (2) The function has the greatest value for $x = pa/(p+q)$. **11.** (5) $\max f = f(-3)$, $\min f = f(-1)$. **14.** Regard the power on the resistance R as a function of the parameter R, find the greatest value of this function on the interval $[0, \infty)$. **16.** The greatest area is equal to $ha/4$. **17.** $t = 3$. **19.** (1) $\max f(x) = f(5)$, $\min f(x) = f(0)$. (2) $\max g(x) = g(0)$, $\min g(x) = g(2)$. (3) $\max S(t) = S(0)$, $\min S(t) = S(-2)$. (4) $\max V(h) = V(1)$, $\min V(h) = V(6)$. **24.** $y'(0) = 0$, $y''(0) = 2$, $y'''(0) = 0$, $y^{(4)}(0) = 24$, $y^{(5)}(0) = 0$. **30.** (1) In the binomial expansion of $(1+x)^n$ set $x = 1$. (2) In the binomial expansion of $(1+x)^n$ set $x = -1$. **31.** $(1.02)^5 < (1.03)^4$. **35.** The points of inflection are $\pm \sqrt{3}/3$. **37.** (1) The point of inflection is $x = -2$. (2) $x = -1$ is a point of minimum, $x = 0$ and $x = 2$ are points of inflection, the straight line $x = 0$ is a vertical tangent to the graph of the function. (6) $x = 5$ is a point of maximum, $x = 1$ is a point of minimum, $x = 5 \pm 2\sqrt{3}$ is a point of inflection. **38.** (1) (a) Yes, it can, (b) Yes, it can, (c) No, it cannot, (d) No, it cannot, (e) No, it cannot. (2) Yes,

it can. (3) It cuts the tangent at the tangent point. **46.** $P(x)$ is divisible by $(x+1)^4$. **55.** We regard $\sqrt{3.14}$ as a value of the function $f(x) = \sqrt{x}$ and consider the error of the measurement of the argument (0.002) to be sufficiently small. Then we can estimate the absolute error of calculation of the value of the function (see Theorem 1 on p. 112) by the expression $0.002/(2\sqrt{3.14}) \cong 0.0006$. Thus we can consider the first three decimal digits in the result 1.7720045 to be valid. **75.** (1) It is easy to guess the root $x = 0$. We show that $f(x)$ is an increasing function and, consequently, has only one root. $f'(x) = e^x(x-1)+1$, $f'(0) = 0$, $f''(x) = xe^x$, $f''(x) < 0$ for $x < 0$ and $f''(x) > 0$ for $x > 0$. Consequently, $f'(x)$ decreases on the left of $x = 0$ and increases on the right of it so that $f'(x) \geqslant 0$ for all x and, hence, $f(x)$ is an increasing function.

Chapter 4

16. $\alpha + \frac{2\pi}{N}k$, $k \in \mathbb{Z}$. **20.** $\varphi = \frac{\pi}{30}t + \varphi_0$ rad or $\varphi = 6t + \varphi_0$ degrees. **22.** (1) 75 m were wound in 5 s, 75 m in 15 s, 0 m in 20 s, and 300 m were wound off the disc in 30 s. (2) 10 s, 100 m, (3) 20 s. (4) 75 m in 5 s, 125 m in 15 s, 200 m in 20 s, 500 m in 30 s. **26.** (1) $\sin\left(\frac{\pi}{2}+2\pi k\right)=1$, $\cos\left(\frac{\pi}{2}+2\pi k\right)=0$. (2) $\sin(\pi+2\pi k)=0$, $\cos(\pi+2\pi k)=-1$. (7) $\sin\left(\frac{5}{6}\pi+2\pi k\right)=\frac{1}{2}$, $\cos\left(\frac{5}{6}\pi+2\pi k\right)=-\frac{\sqrt{3}}{2}$. (8) $\sin\left(-\frac{\pi}{3}+2\pi k\right)=-\frac{\sqrt{3}}{2}$, $\cos\left(-\frac{\pi}{3}+2\pi k\right)=\frac{1}{2}$. **28.** (1) $\frac{\pi}{2}+\pi k$, $k \in \mathbb{Z}$. (2) $-\frac{\pi}{2}+2\pi k$, $k \in \mathbb{Z}$. (3) $\pi+2\pi k$, $k \in \mathbb{Z}$. (5) πk, $k \in \mathbb{Z}$. (6) $\frac{\pi}{2}+\pi k$, $k \in \mathbb{Z}$. (7) $\frac{\pi}{6}+2\pi k$ and $\frac{5}{6}\pi+2\pi k$, $k \in \mathbb{Z}$. (9) $\pm\frac{\pi}{3}+2\pi k$, $k \in \mathbb{Z}$. (11) $-\frac{\pi}{3}+2\pi k$ and $-\frac{2\pi}{3}+2\pi k$, $k \in \mathbb{Z}$. (13) $\pm\frac{5}{6}\pi+2\pi k$, $k \in \mathbb{Z}$. **30.** (1) $\frac{\pi}{6}$ and $\frac{5}{6}\pi$. (2) 0, $\frac{5}{6}\pi$ and $\frac{7}{6}\pi$. (3) $\frac{3}{4}\pi$ and $\frac{5}{4}\pi$. (4) $\frac{\pi}{4}$, $\frac{3}{4}\pi$ and $\frac{\pi}{2}$. **31.** (1) $\frac{\pi}{2}+2\pi k$, $k \in \mathbb{Z}$. (2) $-\frac{\pi}{2}+2\pi k$, $k \in \mathbb{Z}$. (3) $\frac{\pi}{6}+2\pi k$, $\frac{5}{6}\pi+2\pi k$, $-\frac{\pi}{2}+2\pi k$, $k \in \mathbb{Z}$. (4) $\frac{\pi}{2}+\pi k$, $2\pi k$, $k \in \mathbb{Z}$. **32.** (1) $\frac{2}{3}\pi$ and $\frac{\pi}{2}$. (3) 0 and $\frac{2}{3}\pi$. (5) 0, $\frac{\pi}{3}$, $\frac{2}{3}\pi$ and π. **33.** (3) $\frac{7}{12}\pi+2\pi k$, $\frac{13}{12}\pi+2\pi k$, $k \in \mathbb{Z}$. The solution becomes simpler when we note that $\left(x+\frac{\pi}{6}\right)-\left(x-\frac{\pi}{3}\right)=\frac{\pi}{2}$,

hence $\left(x+\frac{\pi}{6}\right)=\frac{\pi}{2}+\left(x-\frac{\pi}{3}\right)$ and $\cos\left(x+\frac{\pi}{6}\right)=-\sin\left(x-\frac{\pi}{3}\right)$. (4) $-\frac{7}{12}\pi+2\pi k$, $-\frac{13}{12}\pi+2\pi k$, $k\in\mathbb{Z}$, note that $\left(x-\frac{\pi}{6}\right)=\left(x+\frac{\pi}{3}\right)-\frac{\pi}{2}$, so that $\cos\left(x-\frac{\pi}{6}\right)=\sin\left(x+\frac{\pi}{3}\right)$. **34.** (1) ω is the angular velocity in radians per second, α_0 is the initial position of the point (for $t=0$). (2) $\frac{2\pi}{\omega}$. **36.** (1) $f(x)=\sin\left(2x+\frac{\pi}{3}\right)=\sin\left(2x+\frac{\pi}{3}+2\pi\right)=\sin\left(2(x+\pi)+\frac{\pi}{3}\right)=f(x+\pi)$. The equality $f(x)=f(x+\pi)$ holds true for any x and, consequently, π is the period of the function $f(x)$. (2) $T=\frac{\pi}{2}$. (3) $T=5\pi$. (4) $T=4\pi$. (5) $T=2\pi$. (6) $T=2\pi$. **37.** $\frac{2}{3}\pi$, $-\frac{4}{3}\pi$. **39.** (1) $f(x)$ is even, $g(x)$ is odd. (2) $f(x)$ is odd, $g(x)$ is even. (3) $g(x)$ is odd, $h(x)$ is even. (4) $f(x)$ is odd, $g(x)$ is even. **43.** $f(0)=0$. **44.** (1) $\sin 38°$ or $\cos 52°$, $-\cos 15°$ or $-\sin 75°$, $-\sin 10°$ or $-\cos 80°$, $-\cos 65°$ or $-\sin 25°$. (4) $\cos 44°$ or $\sin 46°$, $-\sin 26°$ or $-\cos 64°$, $-\sin 13°$ or $-\cos 77°$, $-\sin 51°$ or $-\cos 39°$. (7) $-\sin 0.1\,\pi$, $\cos 0.4\,\pi$, $\sin 0.25\pi$, $\cos 0.15\pi$. **46.** (5) $\cos\left(\frac{\pi}{2}-1\right)$, $-\sin\left(2-\frac{\pi}{2}\right)$, $-\cos\left(2-\frac{\pi}{2}\right)$, $\sin\left(\frac{\pi}{2}-1\right)$. **54.** (1) $T=\pi/2$. (2) $T=2\pi$. (3) $T=3\pi$. (4) $T=\pi/3$. **55.** (1) $\tan\frac{\pi}{6}$. (2) $-\cot 0.1\,\pi$. (3) $\tan 30°$. (4) $\cot 10°$. (5) $-\cot 0.2\pi$. (6) $-\cot 0.2\pi$. (7) $-\cot(\pi-3)$, (8) $-\tan(2\pi-6)$. **57.** (1) $\cos x=-\sqrt{0.91}$, $\tan x=-3/\sqrt{91}$, $\cot x=-\sqrt{91}/3$. (2) x is an angle of the second quadrant, $\sin x=\sqrt{0.84}$, $\tan x=-\sqrt{21}/2$, $\cot x=-2/\sqrt{21}$.

Chapter 5

1. $\sin 15°=\frac{1}{2}\sqrt{2-\sqrt{3}}$, $\cos 15°=\frac{1}{2}\sqrt{2+\sqrt{3}}$. **2.** (1) $\frac{\sqrt{2}}{2}$. (2) -1. **5.** (a) $a\sqrt{1-b^2}+b\sqrt{1-a^2}$, (b) $|a\sqrt{1-b^2}-b\sqrt{1-a^2}|$. **7.** (1) $\sin\left(t+\frac{\pi}{4}\right)$. (2) $\cos\left(2t+\frac{\pi}{4}\right)$. (3) $\sin\left(\frac{t}{2}+\frac{5}{3}\pi\right)$. (4) $\sin\left(t+\frac{11}{6}\pi\right)$. (5) $5\sin(t+\varphi+\pi)$, where $0<\varphi<\pi/2$ and $\sin\varphi=0.8$. (7) $\sqrt{2}\sin 3t$. (8) $\sqrt{2}\cos\left(\frac{1}{3}t+\frac{7}{12}\pi\right)$. **8.** (1) (a) $t=$

$\frac{\pi}{4}+2\pi k$, $k\in\mathbb{Z}$, (b) $t=-\frac{3}{4}\pi+2\pi k$, $k\in\mathbb{Z}$, (c) $t=-\frac{\pi}{4}+\pi k$, $k\in\mathbb{Z}$. (8) (a) $t=-1\frac{3}{4}\pi+6\pi k$, $k\in\mathbb{Z}$, (b) $t=\frac{5}{4}\pi+6\pi k$, $k\in\mathbb{Z}$, (c) $t=-\frac{\pi}{4}+3\pi k$, $k\in\mathbb{Z}$. **9.** (1) 1. (3) $\sqrt{3}\cos\alpha$. **11.** $\tan 15°=2-\sqrt{3}$, $\tan 75°=2+\sqrt{3}$, $\cot\frac{\pi}{12}=\tan 75°$. **18.** (1) $\cos x=0.28$, $\sin x=-0.96$, $\tan x=-3\frac{3}{7}$, $\cot x=-7/24$. **19.** (1) $2\pi k$, $\pm\frac{2}{3}\pi+4\pi k$, $k\in\mathbb{Z}$. (2) $\pi+2\pi k$, $-\frac{\pi}{3}+4\pi k$, $-\frac{5}{3}\pi+4\pi k$, $k\in\mathbb{Z}$. (3) πk, $-\frac{\pi}{6}+2\pi k$, $-\frac{5}{6}\pi+2\pi k$, $k\in\mathbb{Z}$. (4) $\frac{\pi}{2}+\pi k$, $\pm\frac{2}{3}\pi+2\pi k$, $k\in\mathbb{Z}$. (5) πk, $\pm\frac{\pi}{3}+\pi k$, $k\in\mathbb{Z}$. (7) $-\frac{7}{6}\pi+2\pi k$, $k\in\mathbb{Z}$. (9) $-\frac{\pi}{6}+\pi k$, $k\in\mathbb{Z}$. **20.** (1) $\sin x=24/25$, $\cos x=7/25$. **21.** (1) $\frac{1}{2}$ for $x=\frac{\pi}{4}+\pi k$, $k\in\mathbb{Z}$. (3) 2 for $x=\frac{\pi}{4}+\pi k$, $k\in\mathbb{Z}$. **24.** (1) $\sin x=\sqrt{10}/10$, $\cos x=3\sqrt{10}/10$, $\tan x=\frac{1}{3}$, $\cot x=3$. (3) $\sin x=-\sqrt{\frac{17-\sqrt{17}}{34}}$, $\cos x=\sqrt{\frac{17+\sqrt{17}}{34}}$, $\tan x=\frac{\sqrt{18-2\sqrt{17}}}{4}$, $\cot x=-\frac{\sqrt{18+2\sqrt{17}}}{4}$. **26.** (1) $\frac{3}{8}+\frac{1}{2}\cos 2\alpha+\frac{1}{8}\cos 4\alpha$. (3) $\frac{3}{2}-2\sin 2\alpha-\frac{1}{2}\cos 4\alpha$. **29.** (1) $\frac{\pi}{2}k$, $k\in\mathbb{Z}$. (3) πk, $k\in\mathbb{Z}$. (5) $\frac{\pi}{4}+\frac{\pi}{3}k$, $k\in\mathbb{Z}$. **32.** (1) $\frac{\pi}{4}k$, $k\in\mathbb{Z}$. (2) $\frac{\pi}{2}k$, $\frac{\pi}{8}+\frac{\pi}{4}k$, $k\in\mathbb{Z}$. **34.** (1) 0.3948. (2) -2.8577. **35.** $x=2\cos\left(2t+\frac{\pi}{2}-1\right)$, $y=2\cos(2t-1)$. **36.** (1) $\sqrt{2}\sin\left(\frac{1}{2}t+\frac{3}{2}\pi\right)$. **37.** (1) $3\cos(2t+0.8\pi)$. (2) $4\cos(0.5t+1.8\pi)$. (3) $6\cos(3t+0.8\pi)$. (4) $2\cos 4t$. (5) $\frac{1}{2}\cos(0.6t+1.5\pi)$. (6) $\cos 3t$. **42.** (1) $\arcsin 0.4+360°k$, $180°-\arcsin 0.4+360°k$, $k\in\mathbb{Z}$, $\arcsin 0.4+2\pi k$, $\pi-\arcsin 0.4+2\pi k$, $k\in\mathbb{Z}$, $x\cong 23°35'$, $x\cong 156°25'$, $x\cong 0.4115$, $x\cong 2.730$. **47.** (1) $\arctan 2.5+\pi k$, $k\in\mathbb{Z}$, $x\cong 1.2$ on $[0,\pi)$. (2) $-\arctan 4+\pi k$, $k\in\mathbb{Z}$, $x\cong 1.8$ on $[0,\pi)$. (3) $-2\arctan 0.5+2\pi k$, $k\in\mathbb{Z}$, $x\cong 5.4$ on $[0,2\pi)$. (4) $\frac{1}{2}\arctan 2+\frac{\pi}{2}k$, $k\in\mathbb{Z}$, $x\cong 0.55$ on $\left[0,\frac{\pi}{2}\right)$.

(5) $x=\arctan\frac{2}{3}-2+\pi k$, $k\in\mathbb{Z}$, $x\cong 1.7$ on $[0, \pi)$. (6) $-\arctan 1.5+1+\pi k$, $k\in\mathbb{Z}$, $x\cong 0.017$ on $[0, \pi)$. (7) $\frac{\pi}{9}-\frac{1}{3}\arctan\frac{4}{3}+\frac{\pi}{3}k$, $k\in\mathbb{Z}$, $x\cong 0.04$ on $\left[0, \frac{\pi}{3}\right)$, (8) $\frac{\pi}{2}+3\arctan 0.75+\pi k$, $k\in\mathbb{Z}$, $x\cong 3.5$ on $[0, 3\pi)$. 52. (1) On $[0, 2\pi)$, $x_1=\pi/3$, $x_2=\pi$, $x_3=5\pi/3$. (3) On $(-\pi/2, \pi/2)$ $x_1=-\pi/4$, $x_2=-\arctan 0.5$. (8) On $[0, \pi)$ $x_1=\frac{\pi}{8}$, $x_2=\frac{3}{8}\pi$, $x_3=\frac{5}{8}\pi$, $x_4=\frac{7}{8}\pi$, $x_5=0$, $x_6=\frac{\pi}{3}$, $x_7=\frac{2}{3}\pi$. (10) On $[0, 4\pi)$ $x_1=\pi$, $x_2=2\pi$. (12) On $[0, 2\pi)$ $x=\frac{1}{8}\varphi+\frac{\pi}{4}k$, $k=0, 1, \ldots, 7$, $x=\frac{1}{6}\varphi+\frac{\pi}{3}k$, $k=1, 2, 3, 4, 5, 6$, where $\varphi=\arccos 0.6$. (15) On $[0, \pi)$ $x=\frac{\pi}{4}$. 54. (1) $\frac{\pi}{4}+2\pi k<x<\frac{3}{4}\pi+2\pi k$, $\pi+2\pi k<x<\frac{5}{4}\pi+2\pi k$, $-\frac{\pi}{4}+2\pi k<x<2\pi k$. 55. (3) It increases on $\left(2\pi k, \frac{\pi}{2}+2\pi k\right)$ and on $\left(\frac{\pi}{2}+2\pi k, \pi+2\pi k\right)$, decreases on $\left(-\pi+2\pi k, -\frac{\pi}{2}+2\pi k\right)$ and on $\left(-\frac{\pi}{2}+2\pi k, 2\pi k\right)$, $k\in\mathbb{Z}$. 58. $y_{\max}$ for $x=\frac{3}{2}\pi$.

Chapter 6

In exercises 1, 2 and 3 verify the answer by differentiation.

4. (1) $x(t)=\frac{1}{4}t^2-2t+x_0$. (2) $x(t)=-t^2+4t+x_0$.

In exercises 5, 6, 7, 8, 9 and 10 verify the answer by differentiation.

11. (1) $10\frac{2}{3}$. (2) $13\frac{1}{3}$. (3) $\ln 6$. (4) $\ln 5$. (5) $2\frac{2}{3}$. (6) 1.

12. (5) $\ln 2-\frac{1}{2}$. (6) $\sqrt{2}-1$. 14. (8) 5. (9) π. (11) 1. (12) $\frac{\pi(9-4\sqrt{3})}{36}+0.5\ln 1.5$. (13) $\frac{1}{5}e^{\pi}-\frac{2}{5}$. 31. (2) 69.6π. (3) $\frac{\pi}{4}(e^2-1)$. 33. $\frac{8}{3}\pi$. 35. 48. 41. $2\pi rh$. 42. $\frac{\pi}{9}((1-a^4)^{3/2}-1)$. 44. 3π. 52. $x''(t)=\frac{1}{m}\cos t$, $x(t)=-\frac{1}{m}\cos t+c_1t+c_2$, c_1 and

c_2 are constants. 59. (1) 20. (2) 116. 60. $\pi\left(r^2(b-a)-\frac{1}{3}(b^3-a^3)\right)$ for $b>a$. 61. $\pi\rho g\frac{125}{108}R^4$, ρ is the density of water, g is the acceleration of free fall. 63. $\frac{1}{2}\rho g\,dh^2$, ρ is the density of water, g is the acceleration of free fall. 64. $\frac{\gamma mM}{c(c+l)}$, γ is the gravitational constant. 65. $M\omega^2l^2/6$. 66. (3) $F'(x)=2x/\sqrt{1+x^8}$.

Chapter 7

10. (1) $2\left(\cos\frac{\pi}{6}+i\sin\frac{\pi}{6}\right)$. (2) $2\left(\cos\frac{4}{3}\pi+i\sin\frac{4}{3}\pi\right)$. (5) $\sqrt{29}(\cos\varphi+i\sin\varphi)$, where $\varphi=\arccos\frac{5}{\sqrt{29}}$. 12. (1) $-\frac{\pi}{4}+2\pi k$, $k\in\mathbb{Z}$, $1\frac{3}{4}\pi$. (2) $\frac{\pi}{6}+2\pi k$, $k\in\mathbb{Z}$, $\frac{\pi}{6}$. 13. (1) $-i$. (3) $128i$. (5) 256. 15. (1) $\cos(45^\circ+72^\circ\cdot k)+i\sin(45^\circ+72^\circ\cdot k)$, $k=0, 1, 2, 3$, 4.16. (1) $(x-2i)(x+2i)$. (2) $(x+2i)(x-2i)(x-2)(x+2)$. (3) $(x-2i-1)(x+2i+1)$. 18. (1) $x_1=2-i$, $x_2=2+i$.

INDEX

Absolute, error 109, 112
  maximum 80
  value of a complex number 305
Acceleration 34, 276
Addition formulas 166
Amplitude of a function 150
Antiderivative 236
Arc of a curve, length of 270
Arc cosine 186
Arc cotangent 188
Arc sine 183
Arc tangent 187
Area of a curvilinear trapezoid 244
Argument of a complex number 305
Arithmetic value of a root 310
Arrangement 322
Average velocity 33

Bernoulli trial 344
  law of large numbers for 346
Binomial, coefficient 93
  formula 93
  theorem 90, 323
Brownian motion 348
Buffon's needle problem 354

Calculus
  differential 10, 51
  integral 10
Chebyshev's inequality 346
Combination 323
Combinatorics problems 320
Complex number 301
  components of 302
  extraction of a root of 309, 312
  imaginary part of 302
  multiplication of, in trigonometric form 307
  real part of 302
  trigonometric form of 304
Composite function 41
Conjugate complex numbers 303
Cosine of an angle 132
Cosine function 132
  properties of 135
Cotangent of an angle 141
Cotangent function 140
Curvilinear trapezoid, 244
  area of 244

Definite integral 245
  properties of 252
Density of probability 351
Derivative of a function 11, 32
  application of 216
  of a higher order 90, 93
  value of 11
Differential calculus 10, 51
  applications of 70
Differential equation 274, 276
Differential of a function 37
  of a function 41
  of a trigonometric function 212
Differentiation 51
  of an exponential function 63
  of a function defined by an equation 65